

L1 TU
167



ATÍPICOS, CAMBIOS ESTRUCTURALES Y DISCRIMINACIÓN EN SERIES TEMPORALES MULTIVARIANTES

TESIS DOCTORAL

Autor: Pedro Galeano San Miguel.

Director: Daniel Peña Sánchez de Rivera.

UNIVERSIDAD CARLOS III DE MADRID

Departamento de Estadística

Getafe, Marzo de 2004





D. PEDRO GALEANO SAN MIGUEL, con D. N. I. 2243422 W

AUTORIZA:

A que su tesis doctoral con el título: ***"Atípicos, Cambios Estructurales y Discriminación en Series Temporales Multivariantes"*** pueda ser utilizada para fines de investigación por parte de la Universidad Carlos III de Madrid.

Leganés, 17 de mayo de 2004

A handwritten signature in black ink, which appears to read "Pedro Galeano San Miguel".

Fdo.: Pedro Galeano San Miguel

Can't explain?

So BRESAHENTE a com LARGO
(por UNIFORMIDADE)

Julian.

Hubert

[Handwritten signature]

W. G. S. S. S. S. S.

Fuller

Fr: S. V. E. L. L. A

A mi familia y a Conchi.

Agradecimientos.

En la elaboración de esta Tesis, han intervenido, de un modo u otro, muchas personas a quienes quiero mostrar mi gratitud.

En primer lugar, quiero agradecer sinceramente la dedicación y el apoyo del director de esta Tesis, Daniel Peña. Muchas gracias por el esfuerzo, la ayuda, la confianza prestada y el consejo no sólo profesional sino también personal. De Daniel he aprendido muchas cosas acerca del esfuerzo, el trabajo y la paciencia. Gracias a su atención he podido llegar hasta aquí.

Quisiera agradecer al Departamento de Estadística por permitirme realizar esta Tesis doctoral y por la financiación que durante todos estos años me ha permitido su elaboración. En este aspecto quiero también agradecer a todos los profesores de quienes he sido ayudante por permitir de una manera u otra aprender de ellos. El agradecimiento se hace extenso a mis compañeros de los programas de doctorado en Ingeniería Matemática y de Economía con los que he compartido muchos momentos durante estos años.

Quiero mostrar mi gratitud a Daniel Peña, esta vez como investigador principal del proyecto BEC 2000-0167 al Ministerio de Educación y Ciencia por el apoyo financiero ofrecido.

Deseo manifestar mi agradecimiento a Ruey S. Tsay por haberme invitado a visitar la Graduate School of Business de la Universidad de Chicago, lo que constituyó una estancia muy fructífera cuyos resultados están contenidos en esta Tesis y en un próximo artículo. También quiero agradecer a la Fundación Universidad Carlos III por financiar en parte dicho viaje.

También quiero agradecer los consejos de Angeles Carnero, Ana Justel y Pilar Poncela y los momentos que compartimos en los cursos y seminarios de verano de la Universidad de Cádiz que organizaron David Almorza y Alfonso Suárez.

Mi familia siempre me ha ofrecido su apoyo y han mostrado su preocupación por el desarrollo de esta Tesis. Siempre he recibido sus ánimos y su confianza. Quiero dar las gracias también a mis amigos Miguel, Javier, Julio, Pili, Silvia y Pilar por su continuo interés y ánimo. Mi agradecimiento también a Dani y a Montse por hacer de su hogar el mío.

Por último, quiero agradecer a Conchi por su amistad y por ser mi compañera de trabajo

IV

durante todo este tiempo. Hemos pasado por parecidas circunstancias lo que ha hecho que pasáramos mejor los momentos más complicados. Siempre he tenido su apoyo, su interés y su cariño. Gracias por todo.

Índice general

1. Introducción y Resumen	1
2. Algunos problemas relevantes en el análisis de series temporales	5
2.1. Introducción	5
2.2. Modelos	5
2.2.1. Univariantes	5
2.2.2. Multivariantes	8
2.3. Atípicos	9
2.3.1. Atípicos en series univariantes	9
2.3.2. Efectos en la identificación, estimación y predicción	11
2.3.3. Procedimientos de modelización	13
2.3.3.1. Contrastes de la razón de verosimilitudes	13
2.3.3.2. Métodos basados en influencia	14
2.3.3.3. Métodos Robustos	16
2.3.3.4. Métodos Bayesianos	16
2.3.4. Atípicos en series multivariantes	16
2.4. Cambios de varianza	18
2.4.1. Observaciones independientes	18
2.4.2. Observaciones dependientes	19
2.4.3. Modelos de heterocedasticidad condicional	22
2.5. Criterios de selección de modelos	23
2.5.1. Criterios eficientes	24
2.5.1.1. Criterios basados en el error cuadrático medio de predicción	24



2.5.1.2.	Criterios basados en la información de Kullback-Leibler . . .	24
2.5.1.3.	Criterios basados en el análisis espectral	27
2.5.2.	Criterios consistentes	27
2.5.3.	Selección de modelos mediante validación cruzada	29
2.5.4.	Selección de modelos mediante Bootstrap	29
2.5.5.	Algoritmos de selección de modelos ARMA	30
2.5.6.	Criterios para modelos multivariantes	30
2.5.7.	Criterios para modelos TAR	31
2.6.	Conclusiones	31
2.7.	Apéndice	32
3.	Detección de atípicos en series temporales multivariantes mediante Pro-	35
	jection Pursuit	
3.1.	Introducción	35
3.2.	Análisis de atípicos mediante proyecciones	36
3.2.1.	Proyecciones de modelos vectoriales ARMA	36
3.2.2.	Algunas ventajas de buscar atípicos mediante proyecciones	37
3.3.	Direcciones de proyección	41
3.4.	Algoritmo para la detección de atípicos	44
3.4.1.	Cálculo de las direcciones de proyección	45
3.4.2.	Búsqueda de atípicos univariantes	45
3.4.2.1.	Detección de cambios de nivel	47
3.4.2.2.	Algoritmos para la detección del resto de atípicos	49
3.4.3.	Estimación final conjunta	50
3.5.	El caso no estacionario	51
3.6.	Experimentos de Monte Carlo	53
3.6.1.	Valores críticos	53
3.6.2.	Comparación de los estadísticos multivariantes, los estadísticos de proyec- ción y el estadístico cusum para cambio de nivel	55
3.6.3.	Comparación de la potencia de los estadísticos multivariante y de proyección para detectar atípicos	56
3.7.	Ejemplo	56

3.8. Conclusiones	61
3.9. Apéndice	62
4. Cambios de varianza en series temporales multivariantes	69
4.1. Introducción	69
4.2. Contrastes para cambios de varianza en series multivariantes	70
4.3. Procedimientos para la detección de cambios de varianza	73
4.3.1. Procedimiento LRT	74
4.3.2. Procedimiento cusum	76
4.4. Cambios en las varianzas y en las correlaciones	77
4.5. Experimentos de Monte Carlo	79
4.6. Ejemplos	88
4.6.1. El precio de la harina	88
4.6.2. El precio del trigo	91
4.7. Conclusiones	94
4.8. Apéndice	96
5. Criterios de selección de modelos y discriminación cuadrática en series temporales	99
5.1. Introducción	99
5.2. Una familia de criterios de selección de modelos basados en la regla de discriminación cuadrática	100
5.2.1. Una aproximación máximo verosímil	102
5.2.2. Una aproximación Bayesiana	103
5.3. Análisis del término de corrección	104
5.4. Criterios de selección en modelos TAR	107
5.5. Experimentos de Monte Carlo	110
5.5.1. Experimentos para modelos ARMA	110
5.5.2. Experimentos para modelos TAR	114
5.6. Conclusiones	117
5.7. Apéndice	118

6. Futuras líneas de investigación

121

Índice de figuras

2.1. Series con atípicos.	11
2.2. Correlación en los cuadrados de la serie.	19
3.1. Potencia de los estadísticos multivariante y de proyección en función del tamaño del atípico: (a) Potencias Absolutas; (b) Diferencia de Potencias. . .	38
3.2. Tasa de crecimiento del producto nacional bruto de Francia, Italia y España.	59
4.1. Serie bivariante y estadísticos para detección de cambio de varianza.	75
4.2. Índices de precios de la harina en tres ciudades de U.S.A.	88
4.3. Gráfico de puntos de la primera diferencia del logaritmo de cada Índice frente al primer retardo de la misma serie.	90
4.4. Índices de precios del trigo en cinco provincias de Castilla, España	91
4.5. Gráfico de puntos de la primera diferencia del logaritmo de cada Índice frente al primer retardo de la misma serie.	95

Índice de tablas

3.1. Estimación de L cuando $h=T/2$	40
3.2. Modelos para los experimentos.	53
3.3. Valores críticos para los estadísticos considerados para $T=100, 200$ y 500 . . .	54
3.4. Frecuencia de detección de un cambio de nivel para los estadísticos multivariante (J_{\max}), de proyección (Λ_L) y cusum (D_L).	57
3.5. Frecuencia de detección para los estadísticos multivariante y de proyección. .	58
3.6. Atípicos detectados por el algoritmo de proyección.	60
3.7. Estimación de los tamaños de los atípicos detectados por el algoritmo de proyección.	60
3.8. Atípicos detectados por el procedimiento de Tsay, Peña y Pankratz (2000). .	61
4.1. Modelos para el estudio de simulación.	79
4.2. Cuantiles empíricos para los estadísticos Λ_{\max} y Γ_{\max} basados en 10000 series.	80
4.3. Resultados para el modelo 1 y un cambio de varianza.	81
4.4. Resultados para el modelo 2 y un cambio de varianza.	82
4.5. Resultados para el modelo 1 y dos cambios de varianza.	84
4.6. Resultados para el modelo 2 y dos cambios de varianza.	85
4.7. Resultados para el modelo 1 con cambio de varianzas y de correlaciones. . .	86
4.8. Modelos para el estudio de simulación con cambios estructurales.	86
4.9. Resultados para cambios estructurales.	87
4.10. Matriz de parametros AR ($\hat{\Pi}$), matriz de covarianza estimada ($\hat{\Sigma}$) y matriz de correlaciones estimada (R) para los tres modelos ajustados para la serie del precio de la harina. Errores estándar de los coeficientes aparecen bajo paréntesis.	89
4.11. Resumen de los procedimientos LRT y cusum para los datos de la harina. . .	89

4.12. Matriz de parametros AR ($\hat{\Pi}$), matriz de covarianza estimada ($\hat{\Sigma}$) y matriz de correlaciones estimada (\hat{R}) para los tres modelos ajustados para la serie del precio del trigo. Errores estándar de los coeficientes aparecen bajo paréntesis.	92
4.13. Resumen de los procedimientos LRT y cusum para cambios de varianzas para los datos del trigo.	93
4.14. Resumen de los procedimientos LRT y cusum para cambios de varianzas y de correlaciones para los datos de la harina.	94
5.1. Frecuencia de selección del modelo correcto y raíz del error cuadrático medio de predicción para los modelos seleccionados por cada criterio. Si C es un criterio dado, C^* es la versión corregida incluyendo el término de corrección propuesto en esta Tesis.	111
5.2. Media y error estándar de los ordenes seleccionados por los criterios para M1, M2 y M3.	113
5.3. Raíz del error cuadrático medio de predicción para M4.	113
5.4. Frecuencia de selección del orden correcto, raíz del error cuadrático medio de estimación y raíz del error cuadrático medio de predicción suponiendo que el retardo es conocido.	114
5.5. Frecuencia de selección del orden correcto, raíz del error cuadrático medio de estimación y raíz del error cuadrático medio de predicción suponiendo que el retardo es desconocido.	115

Capítulo 1

Introducción y Resumen

Esta Tesis se centra en el análisis de algunas propiedades dinámicas de las series temporales univariantes y multivariantes. Concretamente, se analizan tres problemas habituales en el análisis de una serie temporal:

- (1) Observaciones atípicas, que son aquellas que no siguen la pauta generada por la estructura dinámica de la serie temporal.
- (2) Cambios en las varianzas marginales de las componentes de un vector de series.
- (3) Criterios de selección de modelos.

Estos tres problemas son importantes por las razones siguientes: la presencia de datos atípicos en series temporales es habitual y estos datos pueden afectar mucho el análisis de la serie. Por ejemplo, si se ignora su presencia, se pueden obtener estimadores sesgados de los parámetros, lo que puede conducir a malas predicciones. Un segundo problema es la presencia de cambios en la varianza incondicional, lo que puede afectar a la estimación de la varianza residual, a los intervalos de confianza de predicciones puntuales o a la estimación de las correlaciones al cuadrado de la serie. Los criterios de selección se utilizan para determinar un modelo para ajustar a la serie, pero además son una de las herramientas principales para detectar los efectos comentados anteriormente.

Las aportaciones de esta Tesis contribuyen a la resolución de estos tres problemas. Se propone un método de detección y estimación de datos atípicos en series multivariantes basado en la proyección del vector de series en direcciones adecuadas. El método propuesto tiene la ventaja de no requerir un modelo inicial para la serie. De esta manera se puede limpiar la serie de atípicos y posteriormente, obtener un modelo para estimar conjuntamente los parámetros, tamaños y efectos de los atípicos. Para cambios de varianza, se proponen dos algoritmos para la detección y estimación de cambios en las varianzas marginales de las componentes de una serie multivariante. También se proponen contrastes para identificar cambios de correlaciones entre las componentes del vector de series. Finalmente, se propone un método para seleccionar modelos basado en análisis discriminante que unifica criterios

clásicos y Bayesianos obtenidos mediante diferentes enfoques. Además, este planteamiento permite añadir un término de corrección en estos criterios que, sin modificar sus propiedades asintóticas, mejora su comportamiento en muestras finitas. El método se aplica para seleccionar modelos autorregresivos y de media móvil y un tipo especial de modelos no lineales, los modelos autorregresivos por umbrales.

A continuación se resumen brevemente los contenidos y contribuciones de cada uno de los capítulos de esta Tesis.

El capítulo 2 introduce los tres problemas que se estudian en la Tesis. En primer lugar, se presentan las características básicas de los modelos autorregresivos y de media móvil (ARMA) univariantes y multivariantes, así como los modelos autorregresivos por umbrales (TAR), que se estudian en los siguientes capítulos. Se presenta su definición y las condiciones de estacionaridad e invertibilidad. Para evitar los problemas derivados de la presencia de datos atípicos es importante disponer de técnicas de detección y tratamiento y se revisan los métodos propuestos para tratar con este tipo de datos en modelos ARMA univariantes y multivariantes. A continuación, se analizan cambios en la varianza marginal de una serie temporal. Este tipo de cambio es el equivalente a un cambio de nivel en la media y se observa a menudo en series económicas y medio ambientales. Se revisan los métodos de detección y estimación de estos cambios, que ha sido ampliamente estudiado para series univariantes pero no para series multivariantes. En la práctica rara vez se conocen los ordenes de los modelos ARMA y TAR a ajustar a una serie. Para ello se utilizan criterios de selección de modelos, que minimizan una cierta función objetivo. En este capítulo se revisan algunos de los criterios más utilizados para series univariantes y multivariantes.

En el capítulo 3, se analizan datos atípicos en series temporales multivariantes mediante el uso de proyecciones. Uno de los principales problemas de los métodos existentes es suponer conocido el modelo, lo cual es irreal en la práctica, puesto que la presencia de estos datos hace más difícil la identificación del modelo. El método propuesto permite la detección de atípicos sin necesidad de construir un modelo inicial para la serie multivariante. En este capítulo se obtienen los siguientes resultados:

(1) Se demuestra que un atípico en una serie multivariante produce un atípico en casi toda proyección y que contrastar la presencia del atípico en direcciones de proyección adecuadas puede ser más potente que en la serie multivariante.

(2) Se demuestra que la dirección del atípico es la dirección que produce máxima separación entre los datos limpios y los afectados por el atípico.

(3) Se demuestra que estas direcciones óptimas para detectar atípicos se obtienen mediante la optimización numérica del coeficiente de curtosis de la serie proyectada.

(4) Se propone un método iterativo de Projection Pursuit de búsqueda de estas proyecciones para detectar y estimar varios atípicos basado en dos tipos de estadísticos: los de razón de verosimilitudes y los estadísticos cusum.

(5) Se demuestra que un estadístico cusum es más potente para detectar cambios de nivel

y que es robusto a la presencia de otros atípicos, por lo que se propone detectar primero cambios de nivel y después el resto de datos mediante estadísticos de razón de verosimilitudes.

(6) Se comprueba el buen comportamiento del método mediante un amplio experimento de Monte Carlo, que tiene en cuenta diferentes tipos, tamaño muestral, localización y número de atípicos, además de ilustrar las ventajas del método mediante su aplicación a una serie de datos reales.

En el capítulo 4, se estudia la detección de cambios en las varianzas marginales de las componentes de series temporales multivariantes. En este capítulo se obtienen los siguientes resultados:

(1) Se derivan dos estadísticos para contrastar la presencia de un cambio en las varianzas marginales de las componentes de un vector de series. El primero está basado en la razón de verosimilitudes y el segundo es de tipo cusum, y utiliza las componentes principales de los residuos del modelo.

(2) Se obtiene el estimador máximo verosímil del tamaño del cambio y las distribuciones asintóticas de ambos estadísticos.

(3) Se proponen algoritmos iterativos para cada uno de los estadísticos cuando el número y el lugar de los cambios es desconocido. El primero, para el estadístico de razón de verosimilitudes, se basa en obtener el máximo de los estadísticos en la muestra, contrastar si en el momento en el que se alcanza el estadístico máximo existe un cambio, limpiarlo en caso afirmativo y volver a buscar un nuevo cambio. El segundo, para el estadístico cusum, se basa en buscar el máximo de los estadísticos cusum en toda la muestra, contrastar si en el momento en el que se alcanza el máximo existe un cambio, y si existe, dividir la serie en dos partes, antes y después del momento y buscar nuevos cambios.

(4) Se demuestra mediante un experimento de Monte Carlo que el procedimiento basado en el estadístico cusum es el más potente.

(5) Se derivan el estadístico de razón de verosimilitudes y un estadístico de tipo cusum para contrastar la presencia de cambios en las correlaciones de las componentes del vector de series. Se generalizan los algoritmos descritos para estos estadísticos y se demuestra en un experimento de Monte Carlo que, de nuevo, el algoritmo cusum es más potente.

(6) Se estudia la potencia de los estadísticos propuestos para detectar cambios en los parámetros y se demuestra que estos estadísticos parecen adecuados también para detectar este tipo de cambios.

Como ilustración, se analizan dos series reales de precios y se estudia el funcionamiento en la práctica de los algoritmos propuestos, comprobando que los estadísticos cusum resultan ser más potentes.

En el capítulo 5, se analiza el problema de selección de modelos para series temporales ARMA como un problema de discriminación cuadrática. En este capítulo se obtienen los siguientes resultados:

(1) Se demuestra que este enfoque unifica la derivación de los más conocidos criterios de selección y que también sugiere una definición diferente a la propuesta por Akaike de la esperanza de la verosimilitud. En particular, se demuestra que la regla de discriminación cuadrática mediante un enfoque máximo verosímil proporciona versiones corregidas de los criterios AIC y AICc y con un enfoque Bayesiano proporciona una versión corregida del criterio BIC.

(2) Se obtiene un término de corrección en la definición de los criterios anteriores que depende de la correlación dinámica de la serie. Este factor no modifica el comportamiento de los criterios más habituales en muestras grandes, pero puede producir una gran mejora en muestras pequeñas.

(3) Se extienden estas ideas para modelos autorregresivos por umbrales (TAR).

(4) Se constata en un experimento de Monte Carlo que la corrección introducida mejora el comportamiento en muestras finitas de criterios tales como el AIC, el AIC corregido y el BIC, para modelos ARMA, y de algunos criterios de validación cruzada para modelos TAR.

Por último, en el capítulo 6 se presentan algunas futuras líneas de investigación.

Capítulo 2

Algunos problemas relevantes en el análisis de series temporales

2.1. Introducción

Este capítulo plantea los tres problemas que estudia esta Tesis: análisis de datos atípicos multivariantes, cambios de varianza en series vectoriales ARMA, y criterios de selección de modelos ARMA y TAR univariantes. El resto del capítulo se divide en seis secciones. En la sección 2.2 se presentan las características básicas de los modelos que estudiamos en esta tesis. En la sección 2.3, se ofrece una visión general del problema de los atípicos en series temporales. Se analizan algunas de las consecuencias de la presencia de este tipo de datos en una serie temporal y se revisan los principales métodos de detección y estimación en el caso univariante y multivariante. En la sección 2.4, se presenta el problema de cambios de varianza en series temporales univariantes. En primer lugar se analiza el caso de datos independientes y a continuación, el caso de datos dependientes. En la sección 2.5 se discuten los criterios de selección de modelos en series temporales, analizando los diferentes enfoques para seleccionar los ordenes de los modelos ARMA y TAR. En la sección 2.6 se incluyen algunas conclusiones. Por último, la sección 2.7 es un apéndice donde se derivan los criterios AIC y BIC.

2.2. Modelos

2.2.1. Univariantes

Se supone que una serie temporal, $x = (x_1, \dots, x_T)'$, ha sido generada por la clase de procesos Gaussianos autorregresivos y de media móvil de ordenes p y q , si se verifica:

$$\phi(B)x_t = c + \theta(B)a_t, \quad t = \dots, -1, 0, 1, \dots \quad (2.1)$$

donde B es el operador retardo, $Bx_t = x_{t-1}$, $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ y $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$, son polinomios de grados finitos p y q , c es una constante y a_t es una secuencia de variables aleatorias independientes Gaussianas de media cero y varianza σ_a^2 . Denotaremos a este modelo por $\text{ARMA}(p, q)$. En el caso particular de que $q = 0$, diremos que la serie temporal x_t ha sido generada por un modelo autorregresivo de orden p , $\text{AR}(p)$, y si $p = 0$, diremos que ha sido generada por un modelo de media móvil de orden q , $\text{MA}(q)$.

Los modelos autorregresivos fueron creados con la idea de que el valor actual de la serie pueda ser explicado linealmente en función de los p valores anteriores, donde p es el número de datos necesarios para predecir el valor actual. Estos modelos tienen un número infinito de autocorrelaciones distintas de cero y, como alternativa a ellos, surgieron los modelos de media móvil, donde la correlación entre observaciones distanciadas más de q periodos es 0. Los modelos ARMA aparecen como compromiso entre los dos modelos, siendo principalmente atractivos por ser capaces de modelizar muchas series con un número pequeño de parámetros.

Las condiciones que debe verificar un proceso $\text{ARMA}(p, q)$ como en (2.1) para ser estacionario es que las raíces del polinomio $\phi(z)$ estén fuera del círculo unidad. Esto es, $\phi(z) = 0$ sólo cuando $|z| > 1$. En este caso, el proceso x_t se puede escribir como un proceso lineal,

$$x_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i} = \mu + \psi(B)a_t, \quad (2.2)$$

donde $\psi(B) = \phi(B)^{-1} \theta(B)$. Se dice que (2.2) es la representación de media móvil infinita del proceso x_t . Además, el proceso x_t se dice que es invertible si las raíces del polinomio $\theta(z)$ están fuera del círculo unidad, esto es, $\theta(z) = 0$ sólo cuando $|z| > 1$. En este caso, el proceso x_t se puede escribir como un proceso autorregresivo infinito,

$$x_t = c_\pi + \sum_{i=0}^{\infty} \pi_i x_{t-i} + a_t \implies \pi(B)x_t = c_\pi + a_t, \quad (2.3)$$

donde $\pi(B) = \theta(B)^{-1} \phi(B)$ y se dice que (2.3) es la representación autorregresiva infinita del proceso x_t . En caso de que la serie no sea invertible, también se puede definir (2.3) mediante $\theta(B)\pi(B) = \phi(B)$. Además, siempre se supondrá que los polinomios $\phi(B)$ y $\theta(B)$ del proceso $\text{ARMA}(p, q)$ en (2.1) no tienen raíces comunes.

Un modelo integrado es una generalización de los modelos ARMA para incluir series no estacionarias. Un proceso x_t se dice $\text{ARIMA}(p, d, q)$ si,

$$\phi(B)(1-B)^d x_t = c + \theta(B)a_t, \quad t = \dots, -1, 0, 1, \dots \quad (2.4)$$

es decir, x_t es un $\text{ARIMA}(p, d, q)$ si $(1-B)^d x_t$ es un $\text{ARMA}(p, q)$. En este caso, el valor de la constante es $c = \mu(1 - \phi_1 - \dots - \phi_p)$ donde $E[(1-B)^d x_t] = \mu$. Estos modelos también tienen una representación de media móvil (2.2) definiendo el polinomio $\psi(B)$ mediante $\phi(B)(1-B)^d \psi(B) = \theta(B)$. Se dice que d es el número de raíces unitarias.

A veces, la dependencia del pasado tiende a ocurrir más fuertemente en múltiplos de algún retardo s . Por ejemplo, con datos económicos o medioambientales mensuales donde $s = 12$, o datos trimestrales donde $s = 4$. Por lo tanto, hay que tener en cuenta las fluctuaciones estacionales de estos procesos. El modelo que combina las relaciones estacionales y regulares, es el llamado modelo multiplicativo estacional autorregresivo y de media móvil, $\text{ARIMA}(p, d, q) \times (P, D, Q)_s$ dado por,

$$\phi(B) \Phi(B^s) (1 - B^s)^D (1 - B)^d x_t = c + \Theta(B^s) \theta(B) a_t, \quad t = \dots, -1, 0, 1, \dots \quad (2.5)$$

donde $\phi(B)$ y $\theta(B)$ son polinomios de ordenes p y q , $\Phi(B^s)$ y $\Theta(B^s)$ son polinomios en el operador B^s de ordenes P y Q , y $(1 - B)^d$ y $(1 - B^s)^D$ son las diferencias en B y B^s , respectivamente. Las condiciones de estacionaridad e invertibilidad en el caso $d = D = 0$, se generalizan para los polinomios $\Phi(B^s)$ y $\Theta(B^s)$.

Hasta ahora se han considerado modelos que dependen linealmente de su pasado. A veces, esto no ocurre, por lo que se pueden considerar modelos que sean de la forma,

$$x_t = m(x_{t-1}, x_{t-2}, \dots) + a_t,$$

donde $m(x_{t-1}, x_{t-2}, \dots)$ es una función no lineal de los valores pasados de x_t . En el capítulo 5, se analiza la selección de criterios de un tipo especial de modelos no lineales, los modelos autorregresivos por umbrales (TAR) propuestos por Tong (1983). La idea básica de estos modelos es ajustar modelos $\text{AR}(p)$ localmente. Se supone que se pueden identificar r regiones dependiendo de los valores previos de x_t , de tal manera que la dinámica del sistema cambia en cada una de ellas. Más concretamente, se dice que una serie temporal, $x = (x_1, \dots, x_T)'$ ha sido generada por la clase de modelos TAR si sigue el modelo,

$$x_t = \phi_{j0} + \sum_{i=1}^{p_j} \phi_{ji} x_{t-i} + a_{jt}, \quad \text{si } r_{j-1} \leq x_{t-d} < r_j \quad (2.6)$$

donde a_{jt} , $j = 1, \dots, k$, son series de ruido blanco de media 0 y varianza finita $\sigma_{a_j}^2$, d es un valor entero positivo y $-\infty = r_0 < r_1 < \dots < r_k = \infty$ son llamados los umbrales. La estimación del modelo se realiza por máxima verosimilitud condicional, debido a que la función de verosimilitud depende de los umbrales, que no son parámetros continuos. En el caso de que los modelos (2.6) sean estacionarios, con momentos segundos finitos y la distribución estacionaria de $x = (x_1, \dots, x_T)'$ admita una densidad positiva en todo punto, Chan (1993) demuestra que los estimadores condicionales máximo verosímiles de los parámetros son fuertemente consistentes. Tong (1983) propone un procedimiento iterativo para la identificación y estimación del modelo en (2.6) que depende de un criterio de selección de modelos. Como se ha mencionado, los criterios serán estudiados en una sección próxima por lo que el procedimiento no se detalla aquí.

2.2.2. Multivariantes

Muchas series temporales se analizan mejor cuando se consideran componentes de series temporales multivariantes, X_t , en cuya especificación se incluye la dependencia de cada componente de la serie. Se considera una serie temporal k dimensional $X_t = (X_{1t}, \dots, X_{kt})'$ donde las componentes X_{jt} , $j = 1, \dots, k$ son series univariantes. Una serie multivariante X_t sigue un modelo vectorial autorregresivo y de media móvil de ordenes p y q si sigue el modelo,

$$\Phi(B)X_t = C + \Theta(B)A_t, \quad t = \dots, -1, 0, 1, \dots \quad (2.7)$$

donde $\Phi(B) = 1 - \Phi_1 B - \dots - \Phi_p B^p$ y $\Theta(B) = 1 - \Theta_1 B - \dots - \Theta_q B^q$, son polinomios matriciales de grados finitos p y q sin factores comunes, C es un vector constante de tamaño $k \times 1$ y A_t es una secuencia de variables aleatorias multivariantes independientes Gaussianas de media el vector de ceros y matriz de covarianzas Σ . Denotaremos a este modelo por VARMA(p, q). En el caso particular en que $q = 0$, diremos que la serie temporal X_t ha sido generada por un modelo autorregresivo vectorial de orden p , VAR(p), y si $p = 0$, diremos que la serie temporal X_t ha sido generada por un modelo de media móvil vectorial de orden q , VMA(q).

La serie temporal X_t es estacionaria si, y solo si, $|I - \Phi_1 z - \dots - \Phi_p z^p| \neq 0$ para $|z| \leq 1$, y es invertible si, $|I - \Theta_1 z - \dots - \Theta_q z^q| \neq 0$ para $|z| \leq 1$. La serie X_t tiene una representación autorregresiva, $\Pi(B)X_t = C_\Pi + A_t$, donde $\Pi(B) = \Theta(B)^{-1}\Phi(B) = I - \sum_{i=1}^{\infty} \Pi_i B^i$, $C_\Pi = \Theta(1)^{-1}C$ es un vector de constantes si X_t es invertible y $\Theta(B)\Pi(B) = \Phi(B)$ si no lo es. La representación de media móvil, $X_t = C_\Psi + \Psi(B)E_t$, donde $\Psi(B) = \Phi(B)^{-1}\Theta(B) = I - \sum_{i=1}^{\infty} \Psi_i B^i$, $\Phi(1)C_\Psi = C$ es un vector de constantes si X_t es estacionaria y $\Phi(B)\Psi(B) = \Theta(B)$ si no lo es.

Se dice que una serie no estacionaria de k componentes, está cointegrada si existen menos raíces unitarias que componentes. Si existen h raíces unitarias, entonces hay $k - h$ combinaciones lineales de las componentes del vector de series que son estacionarias. En este caso, la diferenciación de todas las componentes individuales para lograr la estacionaridad lleva a la sobrediferenciación de la serie, lo que conduce a problemas de raíces unitarias en los polinomios de las matrices de media móvil, en cuyo caso obtendremos un modelo no invertible. Engle y Granger (1987) proponen un modelo, que denominan modelo de corrección del error, que no tiene dificultades para la estimación de modelos no invertibles. Para un modelo VARMA(p, q) como en (2.7) donde el número máximo de raíces unitarias de las componentes es 1, el modelo de corrección del error está dado por,

$$\nabla X_t = \alpha \beta X_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \nabla X_{t-i} + A_t - \sum_{i=1}^q \Theta_i A_{t-i},$$

donde α y β son dos matrices de tamaños $k \times (k - h)$ y $(k - h) \times k$ de rango máximo y Φ_i^* son matrices de tamaño $k \times k$ que son funciones de la matrices Φ_i originales. La serie vectorial βX_t contiene a h componentes estacionarios.

2.3. Atípicos

El estudio de datos atípicos tiene una larga historia. Barnett y Lewis (1994) recopilan su estudio en diversas áreas de la Estadística. En esta sección se revisan algunos conceptos básicos del análisis de datos atípicos en series temporales, lo cual incluye sus definiciones, efectos y procedimientos para su detección y tratamiento.

No todos los datos atípicos tienen la misma naturaleza. Por ejemplo, un error de medida, un error de escritura o un error de grabado son errores habituales en el tratamiento de datos. Esto es diferente a determinadas observaciones que en un experimento controlado resultan ser inesperadas, por ejemplo, un pequeño número de observaciones que son heterogéneas con respecto a la gran mayoría y que forman un grupo homogéneo. Existe evidencia, véase Stigler (1986), de que datos procedentes de experimentos muy controlados presentan este tipo de datos heterogéneos. Su estudio puede ser útil para mostrar que una determinada teoría es imperfecta, o para, mejorar el ajuste de modelos que no son capaces de explicar un fenómeno más complejo. Se debe dejar claro que un dato es atípico relativo a un modelo.

2.3.1. Atípicos en series univariantes

Fox (1972) define el atípico innovativo (IO) y el atípico aditivo (AO) en series temporales estacionarias. Supongamos que x_t sigue el modelo ARMA (2.1) y que en lugar de observar x_t , observamos una serie y_t definida como:

$$(IO) \quad y_t = x_t + \psi(B) w I_t^{(h)} \quad (AO) \quad y_t = x_t + w I_t^{(h)},$$

donde $I_t^{(h)}$ es 0 si $t \neq h$ y es 1 si $t = h$ y $\psi(B)$ es la representación de media móvil de la serie x_t . Un IO representa un cambio repentino en el ruido del proceso. Se hace notar que el efecto de este atípico en la serie observada depende de la dinámica del modelo. Un AO representa un cambio externo del valor observado de la serie, como puede ser un error en la recogida del dato. Un AO sólo afecta a una observación de la serie. La serie de residuos es afectada como sigue:

$$(IO) \quad e_t = a_t + w I_t^{(h)} \quad (AO) \quad e_t = a_t + \pi(B) w I_t^{(h)},$$

esto es, en el caso IO, únicamente un residuo es afectado, mientras que un AO afecta a un número de residuos que depende de la dinámica de la serie.

Fox (1972) propone el uso de la razón de verosimilitudes para contrastar la existencia de estos atípicos. Abraham y Box (1979) analizan ambos atípicos desde un punto de vista Bayesiano. Basándose en determinadas probabilidades a priori de que cada observación sea atípica, los autores obtienen las correspondientes probabilidades a posteriori. Chang y Tiao (1983) y Chang, Tiao y Chen (1988) generalizan los resultados de Fox (1972) a modelos ARIMA proponiendo un procedimiento iterativo para detectar innovativos y aditivos y estudian como distinguirlos. Tsay (1986) propone un procedimiento para identificar atípicos de ambos tipos, eliminar sus efectos y especificar un modelo para la serie libre de atípicos.

El primer artículo que introduce cambios de nivel en series temporales no estacionarias es Box y Tiao (1965). Un cambio de nivel (LS) está dado por,

$$y_t = x_t + \frac{w}{1-B} I_t^{(h)} = x_t + w S_t^{(h)},$$

donde $S_t^{(h)}$ es 0 para $t < h$ y es 1 para $t \geq h$. Por lo tanto, la media o el nivel del proceso se modifica hasta el final del periodo observado. Este efecto es diferente del producido por un IO o por un AO, porque un cambio de nivel produce un cambio permanente en la estructura de la serie observada, mientras que un AO o un IO tienen una influencia transitoria. Los residuos están afectados como sigue,

$$e_t = a_t + \frac{\pi(B)}{1-B} w I_t^{(h)} = a_t + \pi(B) w S_t^{(h)}.$$

Box y Tiao (1975) introducen el Análisis de Intervención para modelizar cambios en una serie temporal en un momento temporal conocido. El modelo general está dado por,

$$y_t = x_t + v(B) w I_t^{(h)}, \quad (2.8)$$

donde $v(B)$ es un polinomio en B . Basado en el modelo (2.8), Tsay (1988) considera que puede ser más adecuado el uso de métodos unificados para detectar y estimar atípicos y cambios de estructura en modelos ARIMA, donde los cambios estructurales considerados por Tsay (1988) son cambios de nivel, cambios transitorios y cambios en la varianza de la serie. Un cambio transitorio (TC) está dado por,

$$y_t = x_t + \frac{w}{1-\delta B} I_t^{(h)},$$

donde $0 < \delta < 1$. Cuando $\delta \rightarrow 0$, un TC converge a un AO y cuando $\delta \rightarrow 1$, un TC converge a un LS. Por lo tanto, un TC es un efecto intermedio entre un AO y un LS. También, se hace notar que en algunos modelos un TC puede ser muy próximo a un IO. Los residuos están afectados como sigue,

$$e_t = a_t + \frac{\pi(B)}{1-\delta B} w I_t^{(h)}.$$

Tsay (1988) deriva los estadísticos de la razón de verosimilitudes para contrastar la existencia de un cambio de nivel y de un cambio transitorio.

La figura 2.1 ilustra los cuatro tipos de atípicos en una serie de 100 datos generada a partir de un modelo AR(1). Se comprueba como los tipos IO, AO y TC son atípicos que modifican transitoriamente a una serie estacionaria, mientras que un LS modifica a la serie de manera permanente.

Chen y Tiao (1990) introducen cambios de nivel aleatorios (RLARMA) y estudian la eficiencia con respecto a la estimación del nivel inicial y las predicciones de aproximaciones

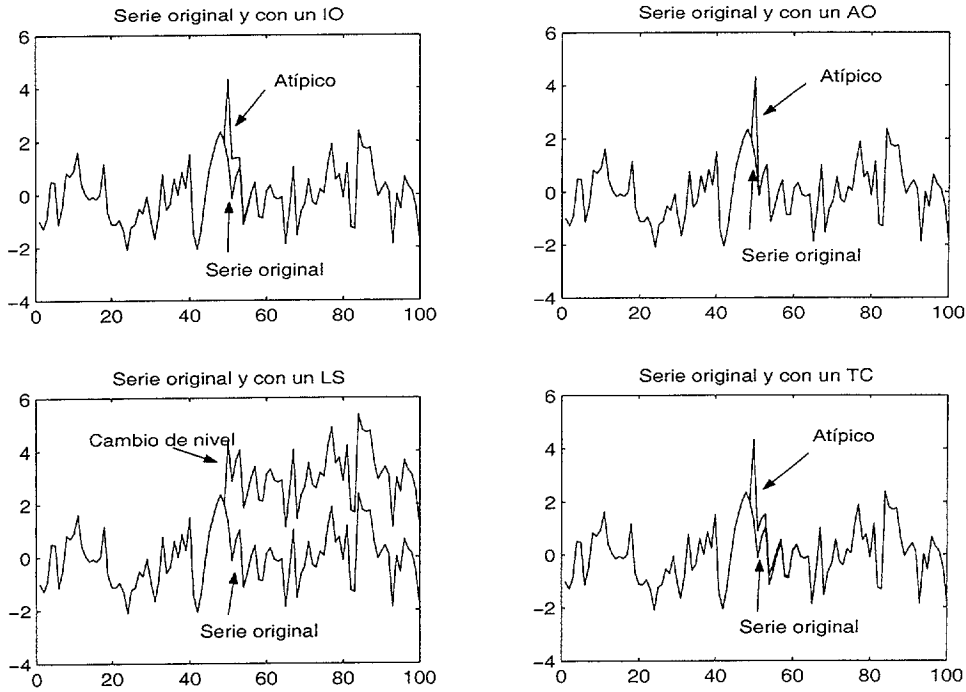


Figura 2.1: Series con atípicos.

ARIMA a series generadas por el modelo RLARMA. El resultado principal es que las aproximaciones ARIMA son inadecuadas para estimar el nivel del proceso. Se propone un procedimiento de detección para tratar diferentes niveles de la serie. En dicho artículo se introduce un nuevo efecto, denominado efecto rampa (RS), dado por,

$$y_t = x_t + \frac{w}{1-B} S_t^{(h)} = x_t + w R_t^{(h)},$$

donde $R_t^{(h)}$ es 0 para $t < h$ y es j para $t \geq h$, donde $t = h + j$. Un efecto rampa produce un cambio en la pendiente de la serie.

También mencionar que Wu, Hosking y Ravishanker (1993) estudian rachas de atípicos consecutivos cuya suma total es cero y que Kaiser y Maravall (2002) estudian atípicos estacionales.

2.3.2. Efectos en la identificación, estimación y predicción

Varios artículos han estudiado los efectos de atípicos y cambios estructurales en áreas relevantes del estudio de series temporales, como son identificación del modelo, estimación y predicción. Los métodos usuales de identificación del modelo ARIMA están basados en la

función de autocorrelación muestral, función de autocorrelación parcial muestral, o mediante criterios de selección de modelos como el criterio de Akaike (AIC) o el criterio Bayesiano (BIC). La presencia de observaciones anómalas puede distorsionar las funciones utilizadas en la identificación del modelo. Por ejemplo, los efectos de atípicos en las funciones de autocorrelación son estudiados en Tsay (1986), Chan (1995) y Trávez y Nievas (1998). Deutsch, Richards y Swain (1990) estudian los efectos de los atípicos en la identificación de modelos ARMA, demostrando que la presencia de un dato atípico en un modelo AR(p) puede llevar a identificar estructura ARMA o MA. Más aun, los ordenes p y q de un modelo ARMA pueden ser mal estimados.

Para ilustrar estos problemas, suponemos que observamos $y_t = x_t + wS_t^{(h)}$, donde el proceso x_t sigue un modelo AR(1) estacionario, $x_t = \phi x_{t-1} + a_t$. Sea $r_y(1)$, la primera autocorrelación del proceso y_t y sea $r_x(1)$, la correspondiente autocorrelación del proceso x_t . Entonces, se puede demostrar que,

$$r_y(1) = \frac{r_x(1) + \frac{\tilde{w}_L(\tilde{y}_{h-1} + \tilde{y}_{h+1} + 2\tilde{y}_{h+2} + \dots + 2\tilde{y}_T) + \tilde{w}_L^2(T-h)}{T}}{1 + \frac{2\tilde{w}_L(\tilde{y}_h + \dots + \tilde{y}_T) + \tilde{w}_L^2(T-h+1)}{T}},$$

donde $\tilde{w}_L = w/\hat{\sigma}_y$ y $\hat{\sigma}_y$ es la varianza muestral de y_t . Si $\tilde{w}_L \rightarrow \infty$ y $T-h$ es lo suficientemente grande, $r_y(1) \rightarrow 1$. Puesto que la primera autocorrelación tiende a 1, el proceso parece tener una raíz unitaria.

Si suponemos que el modelo está bien especificado, los parámetros estimados pueden sufrir de un fuerte sesgo. Algunas referencias a este problema son Chang y Tiao (1983), Tsay (1986) y Chang, Tiao y Chen (1988). Chen y Liu (1993, a) demuestran mediante un estudio de simulación que un AO, LS o un TC pueden producir un fuerte sesgo en los parámetros estimados mientras que un IO tiene un efecto muy pequeño en la estimación. Para ilustrar el efecto en la estimación supongamos que observamos $y_t = x_t + wI_t^{(h)}$, donde el proceso x_t sigue un modelo AR(1) estacionario, $x_t = \phi x_{t-1} + a_t$. El estimador de mínimos cuadrados de ϕ basado en la serie y_t , está dado por,

$$\hat{\phi}_y = \frac{\sum_{t=2}^T y_t y_{t-1}}{\sum_{t=1}^T y_t^2} = \frac{\hat{\phi}_x + \frac{\tilde{w}_A(\tilde{y}_{h-1} + \tilde{y}_{h+1})}{T}}{1 + 2\tilde{w}_A \tilde{y}_h + \frac{\tilde{w}_A^2}{T}},$$

donde $\hat{\phi}_x$ es el estimador mínimo cuadrático del parámetro ϕ basado en x_t , $\tilde{w}_A = w/\hat{\sigma}_y$, $\tilde{y}_t = y_t/\hat{\sigma}_y$. Por lo tanto, si $\tilde{w}_A \rightarrow \infty$, $\hat{\phi}_y \rightarrow 0$. Como ϕ es también la autocorrelación parcial de orden 1, la serie puede ser identificada como un proceso de ruido blanco en lugar de un AR(1). Si la serie observada está afectada por un atípico innovativo, esto es, $y_t =$

$x_t + (1 - \phi B)^{-1} w I_t^{(h)}$, entonces,

$$\hat{\phi}_y = \frac{\sum_{t=2}^T y_t y_{t-1}}{\sum_{t=1}^T y_t^2} = \frac{\hat{\phi}_x + \frac{\tilde{w}_I}{T} \left(\sum_{t=1}^{T-h+1} \tilde{y}_{h-1+t} \phi^t + \sum_{t=0}^{T-h-1} \tilde{y}_{h+1+t} \phi^t \right) + \frac{\tilde{w}_I^2}{T} \frac{\phi}{1-\phi^2}}{1 + \frac{2}{T} \tilde{w}_I \sum_{t=0}^{T-h} \tilde{y}_{h+t} \phi^t + \frac{\tilde{w}_I^2}{T} \frac{1}{1-\phi^2}},$$

y si $\tilde{w}_I^2 \rightarrow \infty$, $\hat{\phi}_y \rightarrow \phi$, y el estimador mínimo cuadrático basado en y_t converge al parámetro autorregresivo.

También existen trabajos que estudian el efecto de atípicos y cambios de nivel en predicción de series temporales. Chen y Tiao (1990) estudian predicciones basadas en aproximaciones ARIMA a modelos RLARMA y concluyen que aproximaciones ARIMA no son robustas para predecir cuando existe un número pequeño de cambios de nivel o cuando la serie tiene una fuerte correlación negativa. Ledolter (1989) y Hotta (1993) estudian el efecto de atípicos aditivos en predicciones. En ambos artículos se concluye que las predicciones puntuales no son excesivamente afectadas por atípicos aditivos pero los intervalos de predicción pueden ser severamente incrementados por los atípicos. El efecto en predicción de LS y TC es considerado en Trávez (1993) quien demuestra que tanto las predicciones puntuales, como los intervalos de predicción están fuertemente afectados. Otro punto de vista es suponer la presencia de atípicos y obtener predicciones. Este planteamiento es adoptado por Hillmer (1984) o Chen and Liu (1993, b), entre otros, quienes demuestran las mejoras de tener en cuenta a los atípicos.

2.3.3. Procedimientos de modelización

2.3.3.1. Contrastes de la razón de verosimilitudes

Estos procedimientos son los más estudiados e implementados en software estadístico. Por ejemplo, los códigos de identificación y estimación de atípicos en SCA y TRAMO están basados en LRT. Se basan en tres pasos fundamentales. Primero, detectar el atípico, segundo, identificar su tipo y tercero, eliminar su efecto en la serie observada para posteriormente especificar un modelo que lo contenga. Suponiendo que los parámetros del modelo son conocidos, los contrastes de la razón de verosimilitudes para la existencia de un IO, AO, LS y TC son,

$$\lambda_{I,h} = \frac{\hat{w}_I}{\sigma_a} \quad \lambda_{A,h} = \frac{\hat{w}_A}{\rho_A \sigma_a} \quad \lambda_{L,h} = \frac{\hat{w}_L}{\rho_L \sigma_a} \quad \lambda_{T,h} = \frac{\hat{w}_T}{\rho_T \sigma_a}, \quad (2.9)$$



respectivamente, donde $\hat{w}_I = e_h$, y,

$$\begin{aligned}\hat{w}_A &= \rho_A^2 (1 - \pi_1 F - \dots - \pi_{T-h} F^{T-h}) e_h & \rho_A^2 &= (1 + \pi_1^2 + \dots + \pi_{T-h}^2)^{-1} \\ \hat{w}_L &= \rho_L^2 (1 - \pi_1^L F - \dots - \pi_{T-h}^L F^{T-h}) e_h & \rho_L^2 &= \left(1 + (\pi_1^L)^2 + \dots + (\pi_{T-h}^L)^2\right)^{-1} \\ \hat{w}_T &= \rho_T^2 (1 - \pi_1^T F - \dots - \pi_{T-h}^T F^{T-h}) e_h & \rho_T^2 &= \left(1 + (\pi_1^T)^2 + \dots + (\pi_{T-h}^T)^2\right)^{-1},\end{aligned}$$

donde F es tal que $Fx_t = x_{t+1}$, $\pi^L(B) = \pi(B)(1-B)^{-1}$ y $\pi^T(B) = \pi(B)(1-\delta B)^{-1}$. Las cantidades \hat{w}_I , \hat{w}_A , \hat{w}_L y \hat{w}_T son las estimaciones del tamaño del atípico bajo la hipótesis de cada tipo.

En la práctica, la localización y el tipo de los atípicos es desconocido. Chang y Tiao (1983) proponen un algoritmo para los casos IO y AO que posteriormente fue extendido para los casos LS y TC por Tsay (1988). El procedimiento tiene diversos pasos. Primero, calcular los estadísticos LRT para todos los tipos y todos los datos de la serie. A continuación, se determina el máximo de los estadísticos (2.9) en valor absoluto. El estadístico que proporcione el valor máximo en el correspondiente tiempo h es el candidato a ser atípico. Si es significativo, su efecto es eliminado de la muestra. Los valores críticos para el máximo en valor absoluto para los LRT se obtienen mediante simulación. Alguna de estas simulaciones para valores críticos aparecen en Chang, Tiao y Cheng (1988), Chen y Liu (1993, *a*) y Sánchez y Peña (2003). Chen y Liu (1993, *a*) y Sánchez y Peña (2003) proponen dos procedimientos alternativos. Aquí mencionamos brevemente el procedimiento de Tsay (1988). Posteriormente, se comentaran los otros dos procedimientos. En primer lugar, ajustar el modelo por máxima verosimilitud bajo la hipótesis de no atípicos en la serie. Entonces, calcular los estadísticos (2.9) para todas las observaciones. Para $1 \leq h \leq T$, obtenemos

$$\mu_h = \max \{|\lambda_{I,h}|, |\lambda_{A,h}|, |\lambda_{L,h}|, |\lambda_{T,h}|\}.$$

Si $\max \mu_h = |\lambda_{I,h}| > C$, donde C es un valor crítico dado, suponemos un IO en $t = h$, y lo mismo ocurre para los otros casos. Cada vez que un atípico es detectado, se elimina su efecto utilizando la serie de residuos. Después de esto, se reestima el modelo y se repite el paso anterior. El algoritmo se itera hasta que no se detectan más atípicos.

2.3.3.2. Métodos basados en influencia

Se denominan observaciones influyentes en regresión, a los puntos con una fuerte influencia en la estimación de los parámetros. Como los atípicos pueden influir en la estimación de los parámetros y en las predicciones, la detección de observaciones influyentes puede ser un arma para la detección de atípicos. Es importante hacer notar que las observaciones influyentes no son siempre atípicas y que los atípicos no son siempre observaciones influyentes. En los modelos de regresión, se definen las medidas de influencia de una observación por el cambio en los parámetros estimados al eliminarla. En series temporales, debida a la correlación entre las observaciones, esto no es tan fácil de generalizar.

Peña (1987, 1990) propone estudiar la influencia de una observación en modelos ARIMA, considerándola como si fuera un dato faltante. Es decir, la observación es reemplazada por su predicción dado el resto de la serie. Se definen dos medidas, una que mide el cambio del vector de parámetros estimados y otra que mide el cambio en la varianza de la serie. Dado un ajuste autorregresivo de orden p de la serie x_t ,

$$x_t = \sum_{i=1}^p \phi_i x_{t-i} + a_t,$$

la medida propuesta para medir el cambio en la estimación de los parámetros de la observación x_h , está dada por,

$$I_\phi(h) = \frac{(\hat{\phi} - \hat{\phi}_{(h)})' \hat{\Sigma}_\phi^{-1} (\hat{\phi} - \hat{\phi}_{(h)})}{p \hat{\sigma}_a^2}, \quad (2.10)$$

donde $\hat{\phi}$ es el estimador máximo verosímil del vector de parámetros $\phi = (\phi_1, \dots, \phi_p)$ para la muestra, $\hat{\phi}_{(h)}$ es el estimador máximo verosímil de ϕ suponiendo que la observación h es faltante y $\hat{\sigma}_a^2 \hat{\Sigma}_\phi$ es la matriz de covarianza de $\hat{\phi}$. La medida de influencia (2.10) se puede escribir en términos de las predicciones basada en ambos vectores de parámetros como sigue,

$$I_\phi(h) = \frac{(\hat{x} - \hat{x}_{(h)})' (\hat{x} - \hat{x}_{(h)})}{p \hat{\sigma}_a^2},$$

donde \hat{x} y $\hat{x}_{(h)}$ son los vectores de predicción para cada vector de parámetros estimados. Se hace notar que esta medida de influencia es válida para series estacionarias y no estacionarias.

Como los atípicos también influyen en la varianza, Peña (1987) propone una medida para la influencia en la varianza dada por,

$$I_{\sigma_a^2}(h) = \frac{\hat{\sigma}_a^2 - \hat{\sigma}_{a(h)}^2}{\hat{\sigma}_{a(h)}^2}, \quad (2.11)$$

donde $\hat{\sigma}_{a(h)}^2$ es la estimación en la varianza con la observación h considerada como faltante. Ambas medidas de influencia pueden ser usadas como estadísticos para observaciones influyentes.

La medida (2.11) es estudiada en Bruce y Martin (1989), Ledolter (1990) y Ljung (1993), entre otros. Peña (1991) extiende los resultados anteriores al caso de modelos de regresión dinámica. Un procedimiento diferente se propone en Chernick, Downing y Pike (1982) y Lefrançois (1991) donde se estudia la influencia de una observación en la función de autocorrelación.

2.3.3.3. Métodos Robustos

Otra manera de tratar con atípicos son los métodos robustos. Estos métodos están diseñados para obtener estimadores no afectados por la presencia de atípicos. En un modelo de regresión, Huber (1973) propone estimadores M, que son el resultado de reemplazar la ecuación de mínimos cuadrados por un sistema equivalente donde la función de minimización es simétrica, convexa, acotada y con valor 0 en el 0. Esta función debe ser robusta a la presencia de atípicos. Denby y Martin (1979) generalizan estimadores M al caso de procesos AR(1). Dada una serie temporal x_t , el estimador GM se obtiene minimizando,

$$\sum_{t=2}^T p(x_t) g(x_t - \phi x_{t-1}),$$

donde $p(\cdot)$ es una función de pesos y $g(\cdot)$ es una función simétrica. En el caso AO, los estimadores GM son más robustos que los estimadores M pero para el caso IO, los estimadores M son más robustos. Martin (1980) y Bustos (1982) extienden estimadores GM a modelos AR(p), Martin (1981) y Lee y Martin (1982) extienden a modelos ARMA y Martin, Samarov y Vandaele (1983) a modelos ARIMA. Otras referencias son Bustos y Yohai (1986), Luceño (1998), García Ben, Martínez y Yohai (2001) y Bianco et al (2001).

2.3.3.4. Métodos Bayesianos

Como se ha mencionado, el primer trabajo Bayesiano que trata de atípicos en series temporales es Abraham y Box (1979). En dicho artículo, se estudia un procedimiento para modelos autorregresivos afectados por innovaciones y aditivos. El principal problema radica en la obtención de las correspondientes probabilidades a posteriori. Los muestreos de Gibbs ayudan a solucionar este problema y varios trabajos utilizan estos métodos para detectar atípicos. McCulloch y Tsay (1993, 1994) presentan procedimientos para estimar los parámetros de modelos autorregresivos y detectar atípicos y para detectar cambios de nivel y cambios en la varianza, respectivamente. Justel, Peña y Tsay (2000) proponen un muestreo de Gibbs modificado para detectar rachas de atípicos.

2.3.4. Atípicos en series multivariantes

Sea X_t un vector de series temporales multivariantes k -dimensional que sigue un proceso ARMA vectorial dado por (2.7). Dada una serie temporal observada $(Y'_1, \dots, Y'_T)'$ donde $Y_t = (Y_{1t}, \dots, Y_{kt})'$, Tsay, Peña y Pankratz (2000) definen cuatro tipos de atípicos univariantes para vectores de series temporales mediante la representación,

$$Y_t = X_t + \alpha(B) w I_t^{(h)}, \quad (2.12)$$

donde $w = (w_1, \dots, w_k)'$, es el tamaño del atípico y X_t sigue un proceso vectorial ARMA. El tipo de atípico se define mediante el polinomio matricial $\alpha(B)$: si $\alpha(B) = \Psi(B)$, se

tiene el atípico innovativo multivariante (MIO), si $\alpha(B) = I$, se tiene el atípico aditivo multivariante (MAO), si $\alpha(B) = (I - B)^{-1}$, se tiene el cambio de nivel multivariante (MLS), y si $\alpha(B) = (I - \delta B)^{-1} I$, se tiene el cambio temporal multivariante (MTC), donde $0 < \delta < 1$ es una constante. Los efectos de estos atípicos sobre los residuos son obtenidos fácilmente cuando se conocen los parámetros del modelo ARMA vectorial para X_t . Entonces, ajustando la serie observada con los verdaderos valores de los parámetros de la serie X_t se obtiene una serie de residuos, E_t , definida por,

$$E_t = \Pi(B)Y_t - C_\Pi, \quad t = 1, \dots, T \quad (2.13)$$

donde $Y_t = X_t$ y $E_t = A_t$ para $t < h$. La relación entre las verdaderas innovaciones de ruido blanco, A_t , y los residuos calculados, E_t , está dada por,

$$E_t = A_t + \Gamma(B)wI_t^{(h)}, \quad (2.14)$$

donde $\Gamma(B) = \Pi(B)\alpha(B)$. Tsay, Peña y Pankratz (2000) demuestran que cuando el modelo es conocido la estimación del tamaño de un atípico multivariante, en tiempo h de tipo i , donde $i = I, A, L, T$, está dada por:

$$w_{i,h} = - \left(\sum_{j=0}^{T-h} \Gamma_j' \Sigma^{-1} \Gamma_j \right)^{-1} \left(\sum_{j=0}^{T-h} \Gamma_j' \Sigma^{-1} E_{h+j} \right), \quad (2.15)$$

donde $\Gamma_0 = -I$. La matriz de covarianzas de esta estimación es $\Sigma_{i,h} = \left(\sum_{j=0}^{T-h} \Gamma_j' \Sigma^{-1} \Gamma_j \right)^{-1}$. De (2.14), tenemos $E_{h+j} = A_{h+j} - \Gamma_j w$, y se puede escribir,

$$w_{i,h} = w - \left(\sum_{j=0}^{T-h} \Gamma_j' \Sigma^{-1} \Gamma_j \right)^{-1} \left(\sum_{j=0}^{T-h} \Gamma_j' \Sigma^{-1} A_{h+j} \right),$$

lo que implica que $\Sigma_{i,h}^{-1/2} w_{i,h}$ está distribuida como $N(\Sigma_{i,h}^{-1/2} w, I)$. Por lo tanto, el estadístico multivariante,

$$J_{i,h} = w_{i,h}' \Sigma_{i,h}^{-1} w_{i,h}, \quad i = I, A, L, T \quad (2.16)$$

se distribuirá como una $\chi_k^2(\eta_i)$ no central con parámetros de no centralidad $\eta_i = w' \Sigma_{i,h}^{-1} w$, para $i = I, A, L, T$. En particular, bajo la hipótesis nula, $H_0 : w = 0$, la distribución de $J_{i,h}$ será χ^2 con k grados de libertad. Un segundo estadístico propuesto por Tsay, Peña y Pankratz (2000) es el estadístico del máximo de los componentes, definido por,

$$C_{i,h} = \max_{1 \leq j \leq k} \frac{|w_{j,i,h}|}{\sqrt{\sigma_{j,i,h}}}, \quad i = I, A, L, T \quad (2.17)$$

donde $w_{j,i,h}$ es el elemento j -ésimo de $w_{i,h}$ y $\sigma_{j,i,h}$ es el elemento j -ésimo de la diagonal principal de la matriz $\Sigma_{i,h}$.

En la práctica, el tiempo h de aparición del atípico y los parámetros del modelo son desconocidos. Entonces, las matrices de parámetros se substituyen por sus estimaciones y se definen los siguientes estadísticos globales:

$$J_{\max}(i, h_i) = \max_{1 \leq h \leq T} J_{i,h}, \quad C_{\max}(i, h_i^*) = \max_{1 \leq h \leq T} C_{i,h}, \quad i = I, A, L, T \quad (2.18)$$

donde se ha llamado h_i y h_i^* a los tiempos en que se alcanzan los máximos de los estadísticos, conjuntos y de componentes. Tsay, Peña y Pankratz (2000) proponen un algoritmo de detección y estimación cuando el número de atípicos es desconocido basado en la obtención del máximo de los estadísticos para cada tipo y suponer un atípico en el caso de que alguno de ellos sea significativo. Si hay varios estadísticos significativos, se selecciona el más significativo de ellos.

2.4. Cambios de varianza

El problema de detectar un cambio repentino en la varianza marginal de una serie temporal univariante ha sido ampliamente estudiado. En esta sección, se revisan los trabajos más relevantes. Se analiza el caso de observaciones independientes y dependientes. Por último, se revisan brevemente los modelos de heterocedasticidad condicional, que modelizan la varianza condicional de la serie temporal igual que los modelos ARMA modelizan la media condicional.

2.4.1. Observaciones independientes

Dada una serie temporal $x = (x_1, \dots, x_T)'$ de observaciones independientes, se dice que x tiene un cambio de varianza en el instante $t = h$, si las observaciones en los bloques de datos x_1, \dots, x_{h-1} y x_h, \dots, x_T tienen varianzas comunes dentro de cada bloque, pero diferentes entre ellos. Un problema importante es detectar la presencia de este cambio y el instante de aparición. Un cambio de varianza en observaciones independientes es que puede introducir correlación en los cuadrados de los datos. Para ilustrar este efecto, se ha generado una serie de 1000 datos Gaussianos independientes de media 0 y varianza 1, y posteriormente se ha definido una nueva serie igual que la primera pero con los últimos 500 datos multiplicados por 2. Se han obtenido las correlaciones de los cuadrados de las dos series que aparecen en la Figura 2.2. Como se puede apreciar un cambio de varianza incluye correlación en la serie de cuadrados.

Hsu, Miller y Wichern (1974) estudian series económicas que presentan evidencias de cambios bruscos en la varianza. Smith (1975) propone un procedimiento Bayesiano para hacer inferencia en el punto de cambio de una distribución, considerando el caso particular de la distribución normal donde puede existir un cambio en la media y la varianza de la serie. Hsu (1977), para el caso de variables normales, propone dos contrastes para cambio de varianza

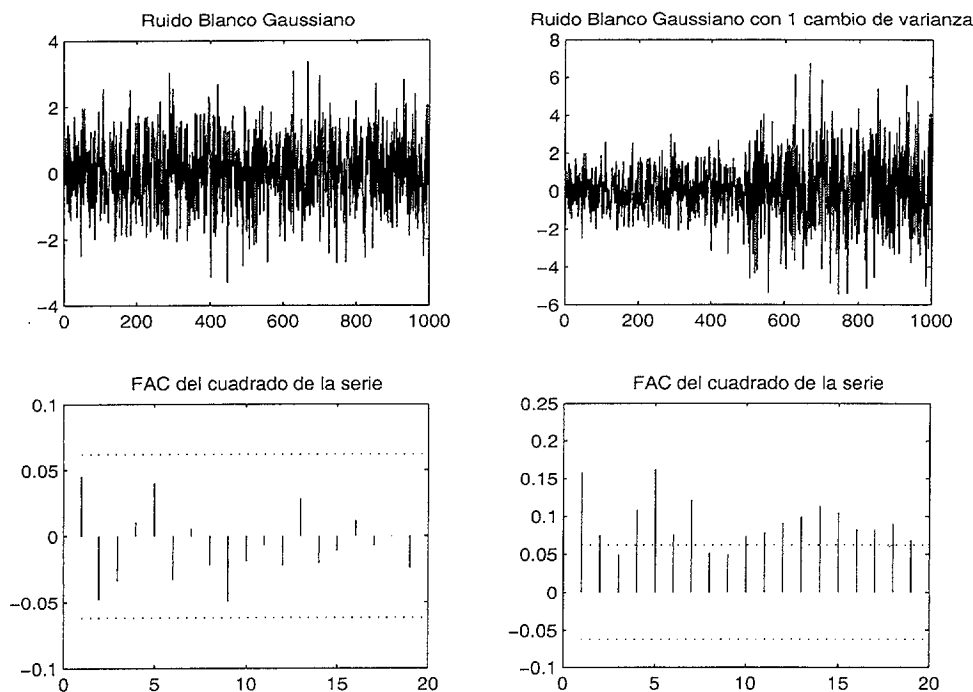


Figura 2.2: Correlación en los cuadrados de la serie.

basados en la suma ponderada de los datos al cuadrado, que es posteriormente extendido por Hsu (1979) al caso de la distribución Gamma. Davis (1979) propone un método robusto para la detección de cambios en la varianza de las innovaciones de una serie temporal. Menzefricke (1981) propone detectar un cambio en la varianza en una secuencia de variables normales mediante el cálculo de la probabilidad a posteriori de que cada una de las observaciones sea un momento de cambio y estudia datos de rendimientos de activos y de procesos de producción.

2.4.2. Observaciones dependientes

Un caso más interesante es el de cambios de varianza en observaciones dependientes. Whichern, Miller y Hsu (1976) estudian estimadores máximo verosímiles del momento de cambio de la varianza en diferentes periodos de tiempo para el caso de un proceso autorregresivo de primer orden. Ali y Giaccotto (1982) estudian contrastes cuando se suponen modelos no paramétricos para series temporales. Abraham y Wei (1984) realizan inferencia en los parámetros de una serie temporal cuando existe la posibilidad de un cambio en su varianza en un punto desconocido. Para ello obtienen la distribución a posteriori de los parámetros y del radio entre varianzas de observaciones antes y después de un determinado punto.

El caso de varios puntos de cambio ha sido también estudiado. Baufays y Rason (1985) estiman las varianzas y los puntos de cambio por máxima verosimilitud y Tsay (1988) propone un algoritmo para detectar atípicos, cambios de nivel y cambios de varianza en modelos ARIMA. Para ello utiliza el contraste de razón de verosimilitudes para detectar un cambio de varianza en $t = h$. Vamos a describir su propuesta.

Sea x_t una serie temporal que sigue un modelo ARIMA como en (2.4). Se supone que en lugar de observar x_t se observa una serie y_t definida como sigue. Sean $S_t^{(h)}$ una función escalón y w una constante que denota el impacto del cambio de varianza. Entonces, las innovaciones que afectan a la serie, e_t , no forma una secuencia de $N(0, \sigma_a^2)$ independientes sino que tienen un cambio en la varianza de la serie en $t = h$, dado por,

$$e_t = a_t + w S_t^{(h)} a_t,$$

de manera que la varianza de e_t cambia de σ_a^2 a $\sigma_a^2(1 + w)^2$ en el punto $t = h$. Sin pérdida de generalidad se asume que $1 + w > 0$ para que w esté bien identificado. La relación entre la serie observada, y_t , y la serie inobservada ARIMA, x_t , vendrá dada por,

$$y_t = x_t + \psi(B) w S_t^{(h)} a_t.$$

Para contrastar la significatividad de un cambio de varianza en $t = h$, se supone que los parámetros del modelo ARIMA en (2.4) son conocidos y se utilizan para calcular los residuos:

$$e_t = y_t - \sum_{i=1}^p \phi_i y_{t-i} - c + \sum_{j=1}^q \theta_j e_{t-j}.$$

El objetivo es contrastar la hipótesis de que estos residuos son homocedásticos, independientes e idénticamente distribuidos, contra la alternativa de que son heterocedásticos. De este modo, se considera la hipótesis nula $H_0 : w = 0$ contra la hipótesis alternativa $H_1 : w \neq 0$. El contraste de razón de verosimilitudes para estas hipótesis es el más potente asintóticamente y está dado por,

$$LR_h = \log \frac{(\hat{\sigma}^2)^T}{(\hat{\sigma}_1^2)^{h-1} (\hat{\sigma}_2^2)^{T-h+1}},$$

donde $\hat{\sigma}^2$, $\hat{\sigma}_1^2$ y $\hat{\sigma}_2^2$ son las varianzas estimadas de los residuos con toda la muestra, con los primeros $h - 1$ residuos y con los últimos $T - h + 1$ residuos, respectivamente. Bajo la hipótesis nula de que no existen cambios de varianza y suponiendo que el modelo es conocido, el estadístico LR_h tiene una distribución asintótica χ^2 con un grado de libertad.

El algoritmo propuesto por Tsay para detectar varios cambios, consiste en obtener el estadístico LR_h para todos los puntos de la muestra, quedarse con el máximo de todos ellos y contrastar si el estadístico asociado a este máximo es mayor que un cierto valor crítico. Si lo es, se considera que en el punto correspondiente hay un cambio de varianza que se estima comparando las varianzas antes y después de ese punto. El cambio se elimina de la serie y se vuelve a iterar el proceso hasta que no se detecten nuevos cambios.

Inclán (1993) propone un procedimiento Bayesiano iterativo basado en cocientes de probabilidades a posteriori para detectar la existencia de varios cambios de la varianza de una serie. Primero, el método se aplica al caso de observaciones independientes, y después se generaliza para modelos autorregresivos. El procedimiento se ilustra, como en varios de los artículos anteriores, con la serie de precios de mercado de IBM.

McCulloch y Tsay (1993) realizan inferencia y predicción sobre cambios en la media y en la varianza de procesos autorregresivos mediante el uso del muestreo de Gibbs. Para ello, generalizan el modelo de Chen y Tiao (1990) para cambios aleatorios en la varianza. Dada una serie x_t que sigue un modelo autorregresivo,

$$x_t = c + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + a_t,$$

donde a_t es una secuencia de variables aleatorias independientes de media 0 y varianzas σ_t^2 tales que,

$$\sigma_t = \begin{cases} \sigma_{t-1} & \text{si } \delta_t = 0 \\ \beta_t \sigma_{t-1} & \text{si } \delta_t = 1 \end{cases}$$

donde δ_t es una secuencia de variables Bernoulli tal que $P(\delta_t = 1) = \epsilon$, ϵ es un valor muy pequeño y β_t es una secuencia de variables aleatorias independientes positivas de alguna distribución conocida. Este modelo es una generalización del propuesto por Tsay (1988) donde están permitidos varios cambios y que permite predecir futuros cambios. Basado en el muestreo de Gibbs, se pueden obtener las probabilidades a posteriori de todos los parámetros del modelo, de donde se puede hacer inferencia acerca de todos ellos. El método es aplicado a una serie de precios mensuales de un cierto tipo de gasolina y los cambios son asociados con diversos hechos que pudieron afectar a dichos precios.

Inclán y Tiao (1994) proponen un procedimiento basado en estadísticos cusum para series de ruido blanco. El estadístico $C_h = \sum_{t=1}^h a_t^2$ es la suma de cuadrados acumulada de una serie de datos incorrelados, a_1, \dots, a_T , de media 0 y varianzas σ_t^2 , $t = 1, \dots, h$. Entonces,

$$D_h = \frac{C_h}{C_T} - \frac{h}{T}, \quad h = 1, \dots, T$$

es la suma de cuadrados acumulada centrada y normalizada. Bajo la hipótesis de homocedasticidad, $\sqrt{T/2}D_h$ se distribuye asintóticamente como un puente Browniano. Inclán y Tiao (1994) proponen un algoritmo basado en el máximo del estadístico $\sqrt{T/2}D_h$ evaluado en todos los puntos de la serie. El máximo de estos estadísticos en valor absoluto tiene una distribución asintótica conocida, ver Billingsley (1968), por lo que el algoritmo comienza buscando el máximo de los estadísticos cusum en valor absoluto. Si este estadístico es mayor que un cierto valor crítico, la serie se divide en dos, se vuelve a calcular los estadísticos cusum en ambas subseries y se busca el máximo. El proceso se itera hasta que no se detectan más cambios. Park, Lee y Jeong (2000) y Lee y Park (2001) extienden el algoritmo de Inclán y Tiao a modelos autorregresivos y de media móvil, respectivamente.

Chen y Gupta (1997) estudian la detección de varios cambios de varianza en una serie de variables Gaussianas independientes. La idea es obtener el criterio BIC para $T + 1$ modelos,

un modelo sin cambios y el resto suponiendo un cambio en cada punto de la serie y obtener el modelo que proporciona el menor valor del BIC. Si se detecta un cambio se procede como en Inclán y Tiao (1994) dividiendo la serie en dos partes, antes y después del cambio y se repite el proceso.

Wang y Zivot (2001) proponen un procedimiento Bayesiano consistente en obtener la probabilidad a posteriori del número de posibles cambios, incluyendo cambios de nivel, de tendencia y de varianza. El lugar de los posibles cambios se determina, como en Chen y Gupta (1997), utilizando el criterio BIC.

Sensier y Dijk (2004) buscan cambios en la varianza incondicional de 214 series macroeconómicas de Estados Unidos en el periodo de 1959 a 1999 mediante un contraste de Wald y concluyen que en más del 80 % de las series existe un cambio en la varianza. Por su parte, Hamori y Tokihisa (1997), Kim, Leybourne y Newbold (2002) y Buseti y Taylor (2003) estudian los efectos de los cambios de varianza en contrastes de raíces unitarias, y proponiendo contrastes adecuados a la presencia de estos cambios.

2.4.3. Modelos de heterocedasticidad condicional

Recientemente, ha existido un gran auge, especialmente en el estudio de series de datos financieros y de alta frecuencia, de modelos que consideran que la presencia de grupos de observaciones consecutivas con diferentes varianzas son debidos a cambios en la varianza condicional de la serie. Los modelos autorregresivos condicionalmente heterocedásticos (ARCH) fueron introducidos por Engle (1982) y posteriormente extendidos a los modelos ARCH generalizados (GARCH) por Bollerslev (1986). Por ejemplo, consideramos el rendimiento o tasa de crecimiento de un activo financiero, x_t , que está definido por, $y_t = (x_t - x_{t-1})/x_{t-1}$, y que típicamente no tiene varianza constante sino que presenta periodos de alta volatilidad. El modelo más simple de heterocedasticidad condicional es el ARCH(1). Supongamos que se ha modelado la media condicional del rendimiento mediante un modelo ARMA, pero que las innovaciones del proceso, a_t , se puede escribir por,

$$\begin{aligned} a_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2, \end{aligned}$$

donde ϵ_t es ruido blanco Gaussiano. El modelo (2.4.3) es un proceso de ruido blanco de tal forma que la varianza condicional depende del valor previo de la innovación. Es fácil ver que la distribución de $a_t|a_{t-1}$ es normal de media 0 y varianza $\alpha_0 + \alpha_1 a_{t-1}^2$, mientras que la distribución de a_t es normal de media 0 y varianza $\alpha_0/(1 - \alpha_1)$. Esto supone que se deben tomar restricciones sobre los parámetros por que si no la varianza puede ser negativa. En este caso, $0 < \alpha_0$ y $\alpha_1 < 1$. Una medida importante para estos modelos es el coeficiente de curtosis dado por,

$$k(a_t) = \frac{E[a_t^4]}{E[a_t^2]^2} = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3,$$

indicando que la distribución marginal de las innovaciones es leptocúrtica, es decir, que tiene colas pesadas. Por otro lado, se puede demostrar que este modelo presenta correlación entre los cuadrados de a_t .

Los modelos ARCH(p) se definen como los ARCH(1), pero con la varianza condicional dada por $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_p a_{t-p}^2$, y muestran propiedades parecidas a las del modelo ARCH(1).

Los modelos GARCH(p, q) son la generalización de los modelos ARCH(p) mediante la inclusión de componentes de media móvil en la ecuación de la varianza condicional. El modelo más estudiado es el GARCH(1,1) dado por,

$$\begin{aligned} a_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{aligned}$$

donde α_1 y β_1 son constantes positivas. Un valor alto de a_{t-1}^2 da lugar a una mayor varianza condicional de la observación a_t , que se atenúan por el coeficiente β_1 , lo que produce rachas de mayor variabilidad. La estimación de estos modelos se realiza mediante máxima verosimilitud condicionada.

Existe un gran número de modelos de heterocedasticidad condicional, por ejemplo, los modelos EGARCH (Nelson, 1991), los modelos de volatilidad condicional de Taylor (1986) y Harvey, Ruiz y Sheppard (1994), los modelos CHARMA de Tsay (1987). Vease Peña, Tiao y Tsay (2000), Carnero, Peña y Ruiz (2001) y Tsay (2002) para la descripción de las características de estos modelos.

Por último, mencionar que también existen trabajos que estudian cambios en la varianza incondicional de una serie que sigue un modelo de heterocedasticidad condicional. Por ejemplo, Kim, Cho y Lee (2000), Lee, Tokutsu y Maekawa (2003) y Sansó, Aragón y Carrión (2003) estudian diferentes contrastes de tipo cusum en series financieras.

2.5. Criterios de selección de modelos

Un punto clave en el análisis de una serie temporal a la que se quiere ajustar un modelo ARMA es determinar sus ordenes. Se han propuesto diferentes ideas para llevar a cabo este objetivo: (1) seleccionar el modelo que predice mejor fuera de la muestra, (2) seleccionar el modelo que minimize una medida de la distancia al modelo verdadero, y (3) seleccionar el modelo con máxima probabilidad a posteriori. Podemos dividir los criterios propuestos en dos clases diferentes, por sus propiedades asintóticas: la clase de los criterios eficientes, que son aquellos que asintóticamente seleccionan el modelo que produce el menor error de predicción, y la clase de los criterios consistentes, que asintóticamente seleccionan el verdadero modelo. Evidentemente, a la vista de las propiedades asintóticas, los puntos de partida de ambos criterios son diferentes. Los criterios eficientes parten de la hipótesis de que el modelo verdadero no tiene por qué pertenecer al conjunto de candidatos, como puede

ocurrir, por ejemplo, en el ajuste autorregresivo de modelos lineales infinitos. Los criterios consistentes parten de que el verdadero modelo está entre los candidatos, por lo que su propiedad asintótica es así justificada.

A continuación se ofrece una revisión de ambas clases de criterios. Hay que notar que en general, los criterios de selección han sido derivado para diferentes modelos estadísticos, por lo que aquí se presentan sus versiones para el caso de series temporales.

2.5.1. Criterios eficientes

2.5.1.1. Criterios basados en el error cuadrático medio de predicción

El primer criterio de selección de modelos fue introducido en Akaike (1969) y es el denominado error de predicción final (FPE). Dada una serie temporal x_1, \dots, x_T , la idea es ajustar modelos autorregresivos de órdenes $1, \dots, p_{\max}$ y seleccionar el modelo que minimice el error de predicción final, que se define como el error cuadrático medio de predicción esperado cuando se ajusta un modelo autorregresivo. El estimador propuesto de este error por Akaike es:

$$FPE(p) = \left(\frac{T + p + 1}{T - p - 1} \right) \hat{\sigma}_p^2, \quad p = 1, \dots, p_{\max}$$

donde $\hat{\sigma}_p^2$ es el estimador máximo verosímil de la varianza residual después del ajuste autorregresivo. Si T es suficientemente grande, el valor de $FPE(p)$ será dominado por el término $\hat{\sigma}_p^2$, que disminuye con el aumento de p . Por lo tanto, el criterio FPE tiende a seleccionar modelos con muchos parámetros. Söderstrom (1977) demuestra que el criterio FPE se puede aplicar a otros modelos de series temporales tales como los modelos ARMA reemplazando $p + 1$ por el número de parámetros del modelo. Bhansali y Downham (1977) sugieren el uso del criterio FPE generalizado, dado por:

$$FPE_\delta(p) = \left(1 + \frac{\delta(p + 1)}{T} \right) \hat{\sigma}_p^2$$

donde δ es una constante positiva. La justificación es que tomando $\delta > 2$, el criterio penaliza más que el FPE por la inclusión de parámetros adicionales. Akaike (1979) analiza la propuesta de Bhansali y Downham, mostrando diversas dificultades.

2.5.1.2. Criterios basados en la información de Kullback-Leibler

Un planteamiento diferente que ha tenido gran impacto en la selección de modelos es el propuesto por Akaike (1973) basado en la teoría de la información. Sea una serie temporal $x = (x_1, \dots, x_T)'$ que ha sido generada por una función de densidad, $p(x|\alpha_0)$, donde α_0 es el vector de parámetros verdaderos. Supongamos un conjunto de modelos ARMA(p, q) que denotamos por $M_{p,q}$ con parámetros $\alpha_{p,q}$ donde $p = 0, \dots, p_{\max}$ y $q = 0, \dots, q_{\max}$, siendo

p_{\max} y q_{\max} dos cotas positivas. La información de Kullback-Leibler para dos funciones de densidad $p(\cdot|\alpha_0)$ y $p(\cdot|\alpha_{p,q})$ está dada por:

$$I(p(\cdot|\alpha_0), p(\cdot|\alpha_{p,q})) = \int p(y|\alpha_0) \log \left(\frac{p(y|\alpha_0)}{p(y|\alpha_{p,q})} \right) dy \geq 0, \quad (2.19)$$

de tal manera que dos densidades estarán cerca cuanto menor sea la información de Kullback-Leibler. Akaike propone seleccionar el modelo que proporcione el valor mínimo de (2.19). Puesto que la distribución $p(y|\alpha_0)$ es desconocida, y para obtener una regla manejable, desarrollando en (2.19) se tiene que,

$$\begin{aligned} I(p(\cdot|\alpha_0), p(\cdot|\alpha_{p,q})) &= \int p(y|\alpha_0) \log p(y|\alpha_0) dy - \int p(y|\alpha_0) \log p(y|\alpha_{p,q}) dy = \\ &= E_y [\log p(y|\alpha_0)] - E_y [\log p(y|\alpha_{p,q})], \end{aligned}$$

donde las esperanzas se toman con respecto a la verdadera distribución $p(\cdot|\alpha_0)$. Puesto que $E_y [\log p(y|\alpha_0)]$ es una constante desconocida, minimizar (2.19) es equivalente a maximizar $E_y [\log p(y|\alpha_{p,q})]$, es decir, seleccionar el modelo que da el máximo del valor promedio de $\log p(y|\alpha_{p,q})$ para todos los valores y de la densidad verdadera $p(y|\alpha_0)$.

En la práctica, dada la serie $x = (x_1, \dots, x_T)'$, la idea más natural es estimar (2.19) mediante:

$$I(p(\cdot|\alpha_0), p(\cdot|\hat{\alpha}_{p,q})) = \int p(y|\alpha_0) \log \left(\frac{p(y|\alpha_0)}{p(y|\hat{\alpha}_{p,q})} \right) dy,$$

donde $\hat{\alpha}_{p,q}$ es el estimador de máxima verosimilitud basado en la serie x . En lugar de esto, Akaike propone obtener el estimador $\hat{\alpha}_{p,q}$ que minimice el valor esperado:

$$E_x [I(p(\cdot|\alpha_0), p(\cdot|\hat{\alpha}_{p,q}))] = E_x \left[E_y \left[\log \left(\frac{p(y|\alpha_0)}{p(y|\hat{\alpha}_{p,q})} \right) \right] \right]. \quad (2.20)$$

Desarrollando (2.20), se tiene que:

$$E_x [I(p(\cdot|\alpha_0), p(\cdot|\hat{\alpha}_{p,q}))] = E_x [E_y [\log p(y|\alpha_0)]] - E_x [E_y [\log p(y|\hat{\alpha}_{p,q})]], \quad (2.21)$$

siendo el primer término de (2.21) un valor constante. Por lo tanto, minimizar (2.20) es equivalente a maximizar $E_x [E_y [\log p(y|\hat{\alpha}_{p,q})]]$. Akaike deriva la siguiente aproximación, como se muestra en el apéndice de este capítulo:

$$E_x [E_y [\log p(y|\hat{\alpha}_{p,q})]] \simeq -\frac{T}{2} \log \hat{\sigma}_{p,q}^2 - (p + q + 1),$$

y que proporciona el criterio de Akaike:

$$AIC(p, q) = T \log \hat{\sigma}_{p,q}^2 + 2(p + q + 1).$$

Otros autores han propuesto derivaciones del criterio AIC bajo hipótesis alternativas, ver Sugiura (1978), Ogata (1980), Chow (1981), Stone (1982), Findley (1985), Bozdogan (1987), Shibata (1989), Cavanaugh (1997) o Burnham y Anderson (2002).

Akaike (1974) aplica el criterio AIC para determinar el orden de ajuste de modelos autorregresivos. En este caso, se verifica que:

$$FPE(p) = \left(\frac{T+p+1}{T-p-1} \right) \hat{\sigma}_p^2 = \left(1 + \frac{2(p+1)}{T} \right) \hat{\sigma}_p^2 + O(T^{-2}),$$

y tomando logaritmos tenemos que asintóticamente, el FPE y el AIC son equivalentes.

Shibata (1976) obtiene la distribución asintótica del orden seleccionado por el criterio de Akaike para modelos autorregresivos y demuestra que dicha distribución depende de los órdenes máximo y mínimo ajustados. A partir de dicha distribución se puede concluir que si el modelo real de la serie se encuentra entre los modelos candidatos, incluso asintóticamente el criterio de Akaike no tiene porque seleccionar el orden verdadero. Hannan (1980) generaliza este resultado para modelos ARMA. Shibata (1980) demuestra que los criterios AIC y FPE son asintóticamente eficientes para procesos autorregresivos infinitos, en el sentido en que los órdenes seleccionados alcanzan asintóticamente el menor error de predicción. Tsay (1984) generaliza estos resultados para el caso de modelos autorregresivos no estacionarios.

Sugiura (1978) obtiene una mejor aproximación para muestras finitas de la distancia de Kullback-Leibler que se deriva en un criterio que denomina AICc y que para modelos ARMA(p, q) viene dado por,

$$AICc(p, q) = \log \hat{\sigma}_{p,q}^2 + \frac{2T(p+q+1)}{T-p-q-2}.$$

Hurvich y Tsai (1989) derivan el criterio de Sugiura para el caso particular de regresión lineal y modelos autorregresivos, demuestran que mejora el comportamiento del AIC en muestras finitas y que es un criterio asintóticamente eficiente. Hurvich, Shumway y Tsai (1990) y Hurvich y Tsai (1991) estudian el sesgo de los criterios AIC y AICc para estimar la información de Kullback-Leibler y proponen un refinamiento basado en un método de Monte Carlo.

Konishi y Kitagawa (1996) derivan un criterio similar al de Akaike, y que denotan por GIC, en el sentido de que es un estimador de la distancia de Kullback-Leibler pero suavizando las hipótesis de partida de Akaike. Es decir, no es necesario la estimación máximo verosímil y la familia de distribuciones paramétricas puede o no contener al verdadero modelo. El criterio obtenido es bastante complicado y depende de funcionales, pero es válido para estimadores de la densidad robustos, Bayesianos y bootstrap.

McQuarrie, Shumway y Tsai (1997) introducen una corrección para modelos de regresión lineal al criterio AICc que denominan AICu que aparentemente funciona mejor en pequeñas muestras en el sentido de que selecciona el verdadero modelo más veces que otros criterios y

que está dado por:

$$AIC_u(p, q) = \log \hat{\sigma}_{p,q}^2 + \frac{2T(p+q+1)}{T-p-q-2} + \log \frac{T}{T-p-q-2}.$$

El inconveniente es que el criterio resultante no es ni consistente, ni eficiente, por lo que no posee propiedades asintóticas conocidas.

Shi y Tsai (1998, 2002) derivan versiones modificadas robustas de los criterios AIC y AICc y proponen un criterio que denominan RIC basada en la verosimilitud de los residuos y que estudia el comportamiento residual después del ajuste de cada modelo candidato.

2.5.1.3. Criterios basados en el análisis espectral

Otro enfoque para seleccionar el orden de un modelo autorregresivo lo propone Parzen (1974, 1975) mediante el uso de un procedimiento basado en la Teoría espectral. Parzen supone que la serie temporal está generada por un autorregresivo infinito y que este proceso puede ser aproximado adecuadamente por un autorregresivo finito. Basado en una medida de distancia entre densidades espectrales, define un criterio que denomina CAT (“criterio para funciones de transferencia autorregresivas”) y que está dado por:

$$CAT(p) = \begin{cases} \frac{1}{T} \sum_{i=1}^p \hat{\sigma}_i^{-2} - \hat{\sigma}_p^{-2} & p = 1, \dots, p_{\max} \\ -(1 + \frac{1}{T}) & p = 0. \end{cases}$$

Posteriormente, Tong (1979) sugiere una modificación del criterio debido a que suele tomar modelos más sencillos de lo deseable y que está dada por:

$$CAT^*(p) = \frac{1}{T} \sum_{i=0}^p \hat{\sigma}_i^{-2} - \hat{\sigma}_p^{-2}. \quad p = 0, \dots, p_{\max}$$

Bhansali (1986) demuestra que los diversos criterios CAT de Parzen y Tong son asintóticamente eficientes en el mismo sentido que el AIC y el FPE en Shibata.

2.5.2. Criterios consistentes

Los criterios consistentes son aquellos que asintóticamente seleccionan el modelo generador de los datos. Por lo tanto, la hipótesis de partida supone que el verdadero generador del conjunto de datos es uno de los modelos candidatos, lo que lleva de manera natural a considerar un punto de vista Bayesiano y seleccionar el modelo con la mayor probabilidad a posteriori dados los datos. Un procedimiento Bayesiano completo a este problema requiere tener en cuenta probabilidades a priori de los modelos y de los parámetros. Por lo tanto,

la derivación de probabilidades a posteriori que requieren integración nos puede llevar a un problema igual de complicado y no comparable desde un punto de vista práctico a la obtención de los valores del criterio de Akaike o similares. Aquí se revisará la obtención de criterios de selección desde un punto de vista Bayesiano, pero no se considera el problema Bayesiano completo y su relación con otros conceptos tales como factores Bayes, métodos MCMC ("Markov Chain Monte Carlo"), promedio Bayesiano, etc... Para un análisis completo del problema véase Kass y Raftery (1995), Kass y Wasserman (1995), Raftery, Madigan y Volinsky (1995) o Chipman, George y McCulloch (2001).

Como en el caso de los criterios eficientes, a continuación se revisan los criterios propuestos en la literatura adaptando las definiciones y desarrollos al caso de series temporales. Para ello y como en el caso de los criterios eficientes, supongamos una serie temporal, $x = (x_1, \dots, x_T)'$ y una serie de modelos $M_{p,q}$ de parámetros $\alpha_{p,q}$ donde $p = 0, \dots, p_{\max}$ y $q = 0, \dots, q_{\max}$, siendo p_{\max} y q_{\max} dos cotas positivas.

Schwarz (1978) con el objetivo de determinar el modelo con mayor probabilidad a posteriori, deriva el criterio Bayesiano (BIC) para el caso de observaciones independientes e idénticamente distribuidas y modelos lineales bajo la hipótesis de que la verosimilitud pertenece a la familia exponencial regular. Para el caso de modelos ARMA(p, q) el criterio BIC se define por:

$$BIC(p, q) = T \log \hat{\sigma}_{p,q}^2 + \log(T) (p + q + 1). \quad (2.22)$$

Se han propuesto diferentes derivaciones del criterio BIC que se pueden encontrar en Chow (1981), Haughton (1988), Raftery, Madigan y Volinsky (1995) y Cavanaugh y Neath (1999). En el apéndice se presenta la propuesta por Chow (1981). Por su parte, Haughton (1988) deriva el criterio BIC mediante el uso del método de aproximación de Laplace de los momentos de distribuciones posteriores (ver Tierney y Kadane, 1986), mientras que Raftery, Madigan y Volinsky (1995) derivan el criterio Bayesiano por el mismo método para promedio de modelos.

Akaike (1981) realiza un análisis crítico de este criterio, argumentando que el uso de (2.22) sólo puede ser justificado tras definir las probabilidades a priori y no omitiéndolas y Chow (1981) demuestra que la aproximación del logaritmo de la probabilidad a posteriori que conlleva el criterio BIC no es buena.

Simultáneamente al trabajo de Schwarz, Akaike (1977) propone maximizar la esperanza del logaritmo de la probabilidad a posteriori del modelo, y que está dado por:

$$BAIC(p, q) = T \log \hat{\sigma}_{p,q}^2 - (T - p - q) \log \left(1 + \frac{p + q}{T} \right) + (p + q) \log(T) + (p + q) \log \left(\frac{\frac{\hat{\sigma}_{p,q}^2}{\hat{\sigma}_x^2} - 1}{p + q} \right),$$

y que, asintóticamente, es similar al criterio propuesto por Schwarz. Además, Akaike justifica su uso porque no requiere probabilidades a priori para cada modelo.

También, mediante un procedimiento no Bayesiano, Rissanen (1978) deriva el criterio (2.22) basándose en el principio de minimizar el número de dígitos binarios requeridos para

describir los datos observados. Hannan y Quinn (1979) buscan un criterio de selección de un modelo autorregresivo de la forma:

$$\log \hat{\sigma}_p^2 + C_T \times p,$$

de manera que el criterio sea consistente y C_T decrezca tan rápido a 0 como sea posible. El criterio obtenido está dado por:

$$HQ(p) = \log \hat{\sigma}_p^2 + 2cp \frac{\log(\log T)}{T}, \quad c > 1$$

pero este criterio tiene un peor comportamiento en muestras finitas que el criterio BIC.

Hannan (1980) demuestra que el orden de un proceso ARMA es estimado consistentemente por los criterios BIC y HQ, pero no por el AIC y determina la distribución asintótica del orden seleccionado por el AIC. Estos resultados son generalizados para procesos autorregresivos no estacionarios por Tsay (1984). Le, Raftery y Martin (1996) proponen una versión del criterio Bayesiano que es robusta frente a atípicos y que está basada en la substitución de la verosimilitud por una verosimilitud robusta.

2.5.3. Selección de modelos mediante validación cruzada

Un procedimiento alternativo de selección de modelos es validación cruzada. Stone (1977) mediante un argumento heurístico deriva la equivalencia asintótica de la selección de un modelo mediante validación cruzada y mediante el criterio de Akaike, lo que fue posteriormente demostrado con mayor detalle por Stoica et al (1986). El método de validación cruzada utiliza los datos pasados y futuros para predecir la observación presente. Un método alternativo es estudiado por Kavalieris (1989), consistente en modificar el método de validación cruzada para que únicamente las observaciones pasadas se utilicen para predecir las observaciones presentes. Kavalieris demuestra que este procedimiento es asintóticamente equivalente al criterio BIC. Cheng y Tong (1992) generalizan estos resultados para el caso de modelos autorregresivos no lineales. Otras referencias son Zhang (1993) y Peña y Sánchez (2004). Por último, De Gooijer (2001) implementa el procedimiento de validación cruzada en modelos TAR y define nuevos criterios de selección basados en los criterios AIC, AICc y AICu para modelos lineales y son comparados por simulación.

2.5.4. Selección de modelos mediante Bootstrap

El bootstrap es un método de Monte Carlo que, en los últimos tiempos, ha gozado de una alta popularidad en aplicaciones estadísticas. Dada una muestra de tamaño T y una serie de modelos, la idea es obtener un número grande B de muestras bootstrap mediante remuestreo, y seleccionar para cada una de ellas, uno de los modelos. El modelo más frecuentemente seleccionado será el elegido. En series temporales, este proceso es algo más

complejo, por la dependencia y posible no estacionaridad de los datos. La idea de utilizar bootstrap para mejorar el comportamiento de los criterios de selección se debe a Efron (1983, 1986) y se discute en Efron y Tibshirani (1993). Algunas referencias básicas de este tipo de procedimientos para la selección de modelos son Shao (1996), Cavanaugh y Shumway (1997), Shibata (1997) e Ishiguro et al (1997).

2.5.5. Algoritmos de selección de modelos ARMA

Algunos autores han estudiado un problema que aparece al ajustar modelos ARMA de órdenes mayores que los verdaderos. Hannan (1980) y Hannan y Rissanen (1982) proponen un procedimiento práctico para selección de órdenes de modelos ARMA mediante la eliminación de los modelos candidatos que sean problemáticos y que además reduce el coste computacional de la estimación de muchos modelos. El procedimiento puede ser adaptado a los criterios AIC o BIC. Hannan y Kavalieris (1984) modifican el método propuesto por Hannan y Rissanen (1982) debido a que dicho método puede sobrestimar alguno de los órdenes. Poskitt (1987), Pötscher (1990) y Pukkila, Koreisha y Kallinen (1990) proponen modificaciones alternativas y demuestran las propiedades de eficiencia y consistencia para los mismos.

2.5.6. Criterios para modelos multivariantes

Los criterios de selección de modelos se generalizan para series multivariantes. De esta manera, si Σ es la matriz de covarianzas de los residuos de dimensión k y $\hat{\Sigma}_{p,q}$ es su estimación máximo verosímil bajo la hipótesis de un modelo ARMA(p, q) vectorial, se definen los siguientes criterios:

$$C(p, q) = \log \left| \hat{\Sigma}_{p,q} \right| + g(T, p, q, k),$$

donde $g(T, p, q, k) = 2\frac{k^2(p+q)}{T}$ para el criterio AIC, $g(T, p, q, k) = \frac{2k^2(p+q)}{T-k(p+q)}$ para el criterio AICc, $g(T, p, q, k) = \log(T)\frac{k^2(p+q)}{T}$ para el criterio BIC y $g(T, p, q, k) = 2\log(\log(T))\frac{k^2(p+q)}{T}$ para el criterio HQ.

Quinn (1980) y Paulsen (1984) demuestran las condiciones de consistencia de los criterios BIC y HQ para series vectoriales estacionarias o con raíces unitarias respectivamente. Lutkepolh (1985) realiza un amplio estudio de simulación comparando los criterios anteriores y otros métodos basados en contrastes para la determinación del orden de una serie autorregresiva vectorial, poniendo de manifiesto las diferencias entre los criterios eficientes y consistentes. Hannan y Deistler (1988) y Poskitt (1989) generalizan el método de Hannan y Kavalieris (1984) y Poskitt (1987) para el caso de vectores ARMA. Koreisha y Pukkila (1993) proponen un método basado en el uso de criterios de selección sobre los residuos para determinar si se acepta un modelo de ruido blanco para los modelos seleccionados previamente en modelos ARMA. Hurvitz y Tsai (1993) derivan el criterio AICc para modelos

autorregresivos.

2.5.7. Criterios para modelos TAR

Por último, se hace mención a los trabajos que consideran criterios de selección para el caso de series que siguen modelos TAR, si bien la definición de los criterios para estos modelos será presentada posteriormente. Tong (1990) proporciona una derivación del criterio de Akaike para este tipo de modelos. Wong y Li (1998) derivan el criterio AICc para estas series, lo comparan con los criterios AIC y BIC en un estudio de simulación y determinan que el criterio AICc es conveniente en muestras pequeñas, si bien en muestras medianas y grandes el BIC funciona mejor. De Gooijer (2001) introduce tres criterios de validación cruzada para estos modelos basados en los criterios AIC, AICc y AICu, los compara mediante simulación y determina que los criterios de validación cruzada funcionan mejor que sus respectivos criterios en versión máximo verosímil. Por último, Kapetanios (2001) proporciona condiciones que debe verificar un criterio para ser consistente para modelos TAR.

2.6. Conclusiones

En este capítulo, se han presentado las características básicas de los modelos ARMA y TAR univariantes y los modelos ARMA vectoriales, que vamos a utilizar en los siguientes capítulos. A continuación se ha presentado el problema de datos atípicos en series temporales, revisando diferentes modelos propuestos y métodos de estimación. Se ha comprobado que este tipo de datos tiene grandes efectos en la identificación, estimación y predicción de una serie, por lo que son necesarios procedimientos para su detección y tratamiento. Se han sido revisado métodos basados en estadísticos de razón de verosimilitudes, en estadísticos cusum, en funcionales de influencia y en métodos Bayesianos y robustos. A continuación, se han analizado cambios de varianza en series temporales con observaciones independientes y dependientes. Se ha comprobado como un sólo cambio en datos independientes puede producir correlaciones significativas en los cuadrados de los datos. Se han revisado métodos de detección y estimación de este tipo de cambios, comprobando que si bien se han propuesto muchos métodos en series univariantes, el caso multivariante no ha sido suficientemente analizado. Por último, se ha estudiado uno de los aspectos más importantes en el análisis de series temporales, como es la selección del modelo a ajustar. Se han revisado diferentes criterios de selección basados en enfoques muy diversos, si bien se puede dar una clasificación de los criterios en dos tipos dependiendo de sus propiedades asintóticas: los criterios eficientes, como el AIC, el AICc o el CAT, y los consistentes, como el BIC y el HQ.

2.7. Apéndice

En este apéndice se derivan los criterios de Akaike (1973) y de Schwarz (1978). En primer lugar, se procede a adaptar la derivación del criterio que propone Akaike para el caso de modelos ARMA(p, q). Para ello, sean x la serie observada, y una serie de observaciones futuras generadas por el verdadero modelo y $\hat{\alpha}_{p,q}$ los estimadores máximo verosimiles de los parámetros bajo la hipótesis de un modelo ARMA(p, q) basados en la serie x . La expansión de segundo orden de $\log p(y|\hat{\alpha}_{p,q})$ alrededor de α_0 está dado por:

$$\begin{aligned} \log p(y|\hat{\alpha}_{p,q}) &\simeq \log p(y|\alpha_0) + \left(\frac{\partial \log p(y|\alpha_0)}{\partial \theta} \right)' (\hat{\alpha}_{p,q} - \alpha_0) + \\ &+ \frac{1}{2} (\hat{\alpha}_{p,q} - \alpha_0)' \left(\frac{\partial^2 \log p(y|\alpha_0)}{\partial \theta \partial \theta'} \right) (\hat{\alpha}_{p,q} - \alpha_0) \end{aligned}$$

Tomando la esperanza de $\log p(y|\hat{\alpha}_{p,q})$ con respecto a y , y teniendo en cuenta que $E_y \left[\left(\frac{\partial \log p(y|\alpha_0)}{\partial \theta} \right) \right] = 0$ y que $E_y \left[\frac{\partial^2 \log p(y|\alpha_0)}{\partial \theta \partial \theta'} \right] = -I(\alpha_0)$, donde $I(\alpha_0)$ es la información de Fisher,

$$E_y [\log p(y|\hat{\alpha}_{p,q})] \simeq E_y [\log p(y|\alpha_0)] - \frac{1}{2} (\hat{\alpha}_{p,q} - \alpha_0)' I(\alpha_0) (\hat{\alpha}_{p,q} - \alpha_0).$$

Tomando la esperanza respecto a x , tenemos:

$$\begin{aligned} E_x [E_y [\log p(y|\hat{\alpha}_{p,q})]] &\simeq E_x [E_y [\log p(y|\alpha_0)]] - \frac{1}{2} E_x [(\hat{\alpha}_{p,q} - \alpha_0)' I(\alpha_0) (\hat{\alpha}_{p,q} - \alpha_0)] = \\ &= E_y [\log p(y|\alpha_0)] - \frac{1}{2} \text{Tr} [I(\alpha_0) \Lambda], \end{aligned}$$

donde $\Lambda = E_x [(\hat{\alpha}_{p,q} - \alpha_0) (\hat{\alpha}_{p,q} - \alpha_0)']$.

En segundo lugar, se obtiene una aproximación del valor de $E_y [\log p(y|\alpha_0)]$. Para ello, se hace notar que:

$$E_y [\log p(y|\alpha_0)] = E_x [\log p(x|\alpha_0)],$$

puesto que x e y son dos series de la misma distribución, $p(y|\alpha_0)$. La expansión de segundo orden de $\log p(x|\alpha_0)$ alrededor de $\hat{\alpha}_{p,q}$ está dado por:

$$\begin{aligned} \log p(x|\alpha_0) &\simeq \log p(x|\hat{\alpha}_{p,q}) + \left(\frac{\partial \log p(x|\hat{\alpha}_{p,q})}{\partial \theta} \right)' (\alpha_0 - \hat{\alpha}_{p,q}) + \\ &+ \frac{1}{2} (\alpha_0 - \hat{\alpha}_{p,q})' \left(\frac{\partial^2 \log p(x|\hat{\alpha}_{p,q})}{\partial \theta \partial \theta'} \right) (\alpha_0 - \hat{\alpha}_{p,q}). \end{aligned}$$

Tomando la esperanza respecto de x y teniendo en cuenta que $\frac{\partial \log p(x|\hat{\alpha}_{p,q})}{\partial \theta} = 0$, y que $E_x \left[\frac{\partial^2 \log p(x|\hat{\alpha}_{p,q})}{\partial \theta \partial \theta'} \right] = -I(\hat{\alpha}_{p,q})$, donde $I(\hat{\alpha}_{p,q})$ es la información de Fisher, se tiene que,

$$\begin{aligned} E_x [\log p(x|\alpha_0)] &\simeq \log p(x|\hat{\alpha}_{p,q}) - \frac{1}{2} E_x [(\alpha_0 - \hat{\alpha}_{p,q})' I(\hat{\alpha}_{p,q}) (\alpha_0 - \hat{\alpha}_{p,q})] = \\ &= \log p(x|\hat{\alpha}_{p,q}) - \frac{1}{2} \text{Tr} [I(\hat{\alpha}_{p,q}) \Lambda]. \end{aligned}$$

Akaike aproxima $I(\hat{\alpha}_{p,q})$ e $I(\alpha_0)$ por Λ^{-1} , ya que asintóticamente son iguales, lo que lleva a la aproximación,

$$E_x [E_y [\log p(y|\hat{\alpha}_{p,q})]] \simeq \log p(x|\hat{\alpha}_{p,q}) - (p+q+1) \simeq -\frac{T}{2} \log \hat{\sigma}_{p,q}^2 - (p+q+1),$$

y que proporciona el criterio de Akaike:

$$AIC(p, q) = T \log \hat{\sigma}_{p,q}^2 + 2(p+q+1).$$

A continuación, se revisa la derivación del criterio Bayesiano propuesta por Chow (1981). Bajo el modelo $M_{p,q}$, x tiene densidad $p(x|\alpha_{p,q}, M_{p,q})$ donde $\alpha_{p,q}$ es el vector de parámetros desconocidos. El procedimiento Bayesiano asigna una distribución de probabilidad a priori a los parámetros del modelo, $p(\alpha_{p,q}|M_{p,q})$ y una probabilidad a priori para cada modelo, $p(M_{p,q})$. La probabilidad de que $M_{p,q}$ sea el modelo generador de los datos, condicionado a la serie x , es la probabilidad a posteriori del modelo,

$$P(M_{p,q}|x) = \frac{P(x|M_{p,q})P(M_{p,q})}{\sum_{p=0}^{p_{\max}} \sum_{q=0}^{q_{\max}} P(x|M_{p,q})P(M_{p,q})},$$

donde,

$$p(x|M_{p,q}) = \int p(x|\alpha_{p,q}, M_{p,q})p(\alpha_{p,q}|M_{p,q})d\alpha_{p,q}.$$

Entonces, la distribución posteriori de $\alpha_{p,q}$ se puede escribir como (ver Zellner, 1987):

$$\begin{aligned} p(\alpha_{p,q} | x, M_{p,q}) &= \frac{p(x | \alpha_{p,q}) p(\alpha_{p,q} | M_{p,q})}{p(x | M_{p,q})} = (2\pi)^{-\frac{p+q+1}{2}} T^{\frac{p+q+1}{2}} |I(\hat{\alpha}_{p,q})|^{\frac{1}{2}} \times \\ &\times \exp \left(-\frac{1}{2} (\alpha_{p,q} - \hat{\alpha}_{p,q})' I_T(x; \hat{\alpha}_{p,q}) (\alpha_{p,q} - \hat{\alpha}_{p,q}) \right) \left(1 + O_p \left(T^{-\frac{1}{2}} \right) \right), \end{aligned}$$

donde $I(\hat{\alpha}_{p,q})$ es la información de Fisher evaluada en los parámetros máximo verosímiles, $\hat{\alpha}_{p,q}$. Tomando logaritmos y evaluando esta expresión en $\hat{\alpha}_{p,q}$, se tiene que,

$$\begin{aligned} \log p(x | M_{p,q}) &= \log p(x | \hat{\alpha}_{p,q}) + \log p(\hat{\alpha}_{p,q} | M_{p,q}) + \\ &+ \frac{p+q+1}{2} \log 2\pi - \frac{p+q+1}{2} \log T + \frac{1}{2} \log |I(\hat{\alpha}_{p,q})| + O_p \left(T^{-\frac{1}{2}} \right), \end{aligned}$$

y la probabilidad a posteriori de cada modelo será proporcional a,

$$P(M_{p,q}|x) \propto \log(M_{p,q}) + \log p(x | M_{p,q}) = \log(M_{p,q}) + \log p(x | \hat{\alpha}_{p,q}) + \\ + \log p(\hat{\alpha}_{p,q} | M_{p,q}) - \frac{1}{2}(p+q+1) \log \frac{T}{2\pi} + \frac{1}{2} \log |I_T(x; \hat{\alpha}_{p,q})| + O_p\left(T^{-\frac{1}{2}}\right),$$

lo que deriva el criterio Bayesiano tras eliminar constantes e ignorar las probabilidades a priori:

$$BIC(p, q) = T \log \hat{\sigma}_{p,q}^2 + \log(T)(p+q+1).$$

Capítulo 3

Detección de atípicos en series temporales multivariantes mediante Projection Pursuit

3.1. Introducción

Construir modelos para series temporales multivariantes puede ser una labor complicada, especialmente cuando la serie está afectada por atípicos. En este capítulo, se propone un método para identificarlos que no requiere una especificación inicial del modelo multivariante. El método está basado en la detección de atípicos univariantes en las proyecciones del vector de series temporales en algunas direcciones útiles. Un atípico multivariante produce al menos un atípico univariante en casi toda serie proyectada y la detección de estos atípicos univariantes conduce a identificar los correspondientes multivariantes. Se demuestra que es posible identificar los atípicos multivariantes mediante estadísticos univariantes en ciertas direcciones óptimas, construidas maximizando y minimizando el coeficiente de curtosis de la serie proyectada. Se propone un algoritmo iterativo basado en proyecciones para limpiar la serie de atípicos.

Este capítulo está organizado como sigue. En la sección 3.2, se introduce la notación y se discuten las propiedades de los atípicos univariantes obtenidos mediante proyecciones del vector de series temporales y se analizan algunas de las posibles ventajas de utilizar proyecciones para detectar atípicos. En la sección 3.3, se demuestra que las direcciones óptimas para identificar atípicos pueden ser obtenidas mediante la maximización y minimización del coeficiente de curtosis de la serie proyectada. En la sección 3.4, se propone un algoritmo de detección basado en proyecciones. En la sección 3.5, se generaliza el procedimiento para series no estacionarias. En la sección 3.6, se analiza el comportamiento del procedimiento propuesto mediante un estudio de Monte Carlo. En la sección 3.7, el método se ilustra con una serie de datos reales. Por último, la sección 3.8 concluye el capítulo.

3.2. Análisis de atípicos mediante proyecciones

En esta sección se explora la utilidad de proyectar el vector de series temporales para detectar atípicos. Primero se estudia la relación entre los modelos univariantes proyectados y el multivariante y, a continuación, se analizan las posibles ventajas de buscar atípicos en la serie proyectada.

3.2.1. Proyecciones de modelos vectoriales ARMA

Se supone un vector de series $X_t = (X_{1t}, \dots, X_{kt})'$, $t = 1, \dots, T$ de k componentes que sigue un proceso ARMA vectorial dado por (2.7). Se supone que en lugar de observar la serie X_t , se observa la serie Y_t dada en (2.12), donde el polinomio matricial $\alpha(B)$ depende del tipo de atípico multivariante. Es bien conocido que si proyectamos el modelo ARMA vectorial en (2.7) en la dirección v , se obtiene un modelo ARMA univariante (ver, por ejemplo, Lütkepohl, 1993). Sea $x_t = v'X_t$. Si X_t es un vector ARMA(p, q), entonces x_t es ARMA(p^*, q^*), con $p^* \leq kp$ y $q^* \leq (k-1)p + q$. En particular, si X_t es un vector MA(q), entonces x_t es un MA(q^*), con $q^* \leq q$; y si X_t es un vector AR(p), entonces x_t es un ARMA(p^*, q^*), con $p^* \leq kp$ y $q^* \leq (k-1)p$. En general, el modelo de la serie univariante es,

$$\phi(B)x_t = c + \theta(B)a_t, \quad (3.1)$$

donde $\phi(B) = |\Phi(B)|$, $c = v'\Phi(1)^*C$ y $v'\Omega(B)A_t = \theta(B)a_t$, donde $\Phi(1)^*$ es el adjunto de $\Phi(1)$, $\Omega(B) = \Phi(B)^*\Theta(B)$ y a_t es un proceso de ruido blanco escalar con varianza σ_a^2 . Los valores para $\theta(B)$ y σ_a^2 se pueden obtener mediante el algoritmo propuesto en Maravall y Mathis (1994), que siempre proporciona una representación invertible del proceso univariante. La representación autorregresiva del modelo univariante (3.1) es $\pi(B)x_t = c_\pi + a_t$, donde $c_\pi = \theta(1)^{-1}c$ y $\pi(B) = \theta(B)^{-1}\phi(B) = 1 - \sum_{i=1}^{\infty} \pi_i B^i$ y su representación de media móvil es $x_t = c_\psi + \psi(B)a_t$, donde $c_\psi = \phi(1)^{-1}c$ y $\psi(B) = \phi(B)^{-1}\theta(B) = 1 + \sum_{i=1}^{\infty} \psi_i B^i$.

Cuando la serie observada Y_t está afectada por un atípico, como en (2.12), la serie proyectada $y_t = v'Y_t$ verifica $y_t = x_t + v'\alpha(B)wI_t^{(h)}$. Si el atípico multivariante es aditivo, la serie proyectada es $y_t = x_t + \beta I_t^{(h)}$, y tiene un atípico aditivo de tamaño $\beta = v'w$ en $t = h$. De la misma manera se comprueba que la proyección de un vector de series temporales con un cambio de nivel multivariante de tamaño w produce una serie univariante con un cambio de nivel de tamaño $\beta = v'w$. El mismo resultado se aplica para cambios transitorios. Por lo tanto, para estos tipos de atípicos las siguientes hipótesis son equivalentes:

$$\begin{aligned} H_0 : w = 0 & \Leftrightarrow H_0^* : \beta = 0 \\ H_A : w \neq 0 & \Leftrightarrow H_A^* : \beta \neq 0 \end{aligned} \quad \forall v \in S^k - \{v \perp w\}$$

debido a que, $H_0 = \{\cap H_0^* : v \in S^k - \{v \perp w\}\}$, donde $S^k = \{v \in R^k : v'v = 1\}$.

Los atípicos innovativos multivariantes producen una racha de atípicos consecutivos de tamaños $v'w$, $v'\Psi_1 w, \dots, v'\Psi_{T-h} w$. Suponiendo que h no es muy próximo a T y debido a

que $\Psi_j \rightarrow 0$, el tamaño de esta racha de atípicos tiende a 0. En el caso particular de que $v'\Psi_i w = \psi_i v'w$, $\forall i = 1, \dots, T-h$, entonces y_t tiene un IO en $t = h$ de tamaño $\beta = v'w$. Pero si $v'\Psi_i w = 0$, $\forall i = 1, \dots, T-h$, entonces y_t tiene un atípico aditivo en $t = h$ de tamaño w , y si $v'\Psi_j w = v'w$, $\forall j = 0, \dots, T-h$, entonces y_t tiene un cambio de nivel en $t = h$ de tamaño $\beta = v'w$. Por lo tanto, la serie univariante y_t obtenida por la proyección puede estar afectada por la presencia de un atípico aditivo, una racha de atípicos o un cambio de nivel.

3.2.2. Algunas ventajas de buscar atípicos mediante proyecciones

La primera ventaja del uso de proyecciones de las series multivariantes para la búsqueda de atípicos es la simplicidad. Utilizando la serie univariante no necesitamos construir un modelo multivariante completo para la serie. La segunda, cuando los parámetros son conocidos, una dirección de proyección conveniente producirá estadísticos que pueden ser más potentes que los multivariantes. La tercera, que cuando los parámetros no son conocidos el sesgo en su estimación es menor en los modelos de proyección que en los multivariantes.

Para ilustrar la segunda ventaja, consideremos una serie Y_t que sigue un modelo ARMA vectorial en (2.7) afectada por un atípico multivariante del tipo MAO, MLS o MTC, en tiempo $t = h$. Sea la matriz V de tamaño $k \times k$ cuya primera columna es $w/\|w\|$, mientras que el resto son $k-1$ vectores unitarios ortogonales a la dirección del atípico. Entonces la serie multivariante $V'Y_t$ estará afectada por un atípico de tamaño $(\|w\|, 0, \dots, 0)'$ que únicamente afecta a la primera componente. Puesto que el estadístico multivariante $J_{i,h}$ es invariante ante transformaciones lineales, su valor coincidirá para las series Y_t y $V'Y_t$. Esto demuestra que toda la información referente a la detección del atípico se encuentra en la primera componente de $V'Y_t$, que es la proyección del vector de series en la dirección del atípico, mientras que el resto de componentes son irrelevantes para detectarlo.

Para una mejor ilustración, se considera el estadístico $J_{i,h}$ $i = I, A, L, T$, que se distribuye como una $\chi_k^2(\eta_i)$ no central con parámetro de no centralidad $\eta_i = w'\Sigma_{i,h}^{-1}w$. La potencia del estadístico $J_{i,h}$ está dada por $Pow(M) = \Pr(J_{i,h} > \chi_{k,\alpha}^2)$, donde $\chi_{k,\alpha}^2$ es el percentil α de la distribución chi cuadrado con k grados de libertad. Si se proyecta la serie Y_t en la dirección v , se obtiene una serie y_t afectada por un atípico y el estadístico $J_{i,h}$ en el caso univariante, $j_{i,h}$, se distribuye como una $\chi_1^2(\tilde{\eta}_i)$ no centrada con parámetro de no centralidad $\tilde{\eta}_i = \beta^2/\sigma_{i,h}^2$, donde $\beta = v'w$, $\sigma_{i,h}^2 = Var(\beta_{i,h})$ y $\beta_{i,h}$ es el estimador de β . La potencia del estadístico univariante $j_{i,h}$ es $Pow(U) = \Pr(j_{i,h} > \chi_{1,\alpha}^2)$. Supongamos para simplificar que la serie es ruido blanco y, sin pérdida de generalidad porque el procedimiento que se propone es afín equivariante, que $\Sigma = I$. Sea $v = \xi w$, donde ξ es una constante normalizadora tal que $v'\Sigma_Y v = 1$, donde Σ_Y es la matriz de covarianza poblacional de Y_t . Entonces, es fácil ver que para todo w , $\eta_i = \tilde{\eta}_i = w'w$, para $i = I, A$, $\eta_L = \tilde{\eta}_L = (T-h+1)w'w$ y $\eta_T = \tilde{\eta}_T = \frac{1-\delta^2(T-h+1)}{1-\delta^2}w'w$. Las potencias, $Pow(U)$ y $Pow(M)$, y sus diferencias $Pow(U) - Pow(M)$ para el caso de un MIO y de un MAO se muestran en la Figura 3.1 para diferentes valores de $w'w$. La figura muestra que si suponemos conocido

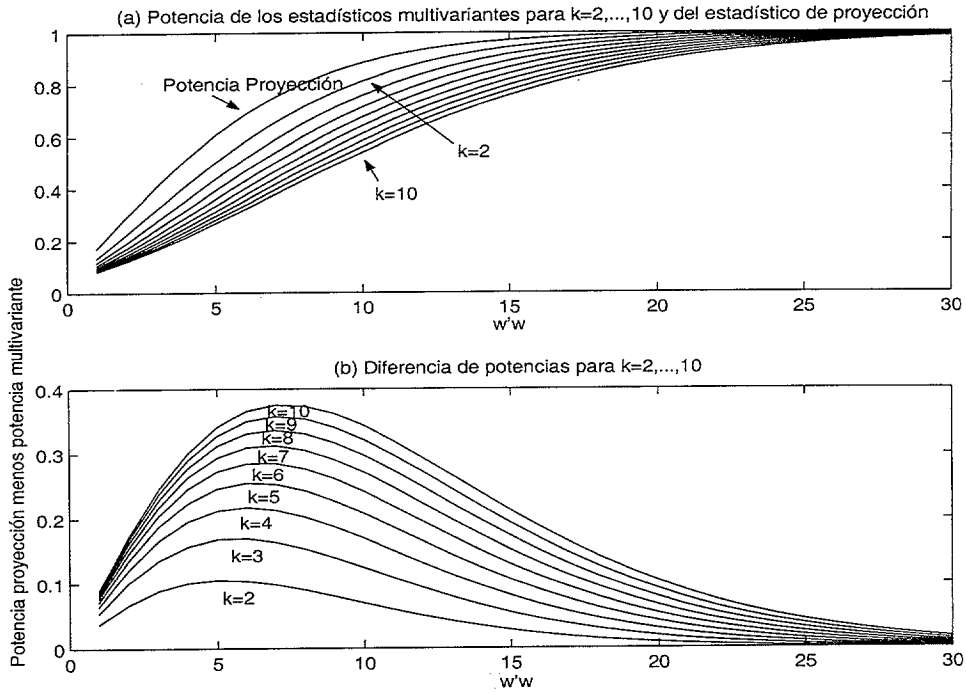


Figura 3.1: Potencia de los estadísticos multivariante y de proyección en función del tamaño del atípico: (a) Potencias Absolutas; (b) Diferencia de Potencias.

el modelo y el tamaño del atípico, cuanto mayor sea el número de componentes, mayor es la ventaja del estadístico de proyección con respecto al multivariante. Cuando el tamaño del atípico crece ambos estadísticos tienen potencias cercanas a uno y por lo tanto sus diferencias tienden a cero. En este análisis hemos supuesto que el vector de series es de ruido blanco. Para procesos vectoriales ARMA la diferencia de potencia entre los estadísticos multivariante y de proyección dependen del modelo. En la sección 3.4, se comparan las potencias relativas de ambos tipos de estadísticos mediante un estudio de simulación.

Vamos a ver que cuando la serie está afectada por un atípico, las estimaciones de los parámetros y de la matriz de covarianzas estarán sesgadas, y que este sesgo, puede reducir fuertemente la potencia del estadístico multivariante. Este sesgo además aumenta con el número de componentes. Vamos a centrarnos en el sesgo en la matriz de covarianzas residual. Como antes, se considera el caso de una serie de ruido blanco de k componentes, Y_t , afectada por un atípico en $t = h$, $Y_t = E_t = A_t + \Gamma(B)wI_t^{(h)}$, donde $w = m1_k$, m es una constante y 1_k es un vector k -dimensional con unos en todas sus componentes, de manera que todas las componentes del vector de series temporales están afectadas igual por el atípico. En el apéndice de este capítulo se demuestra el Lema y Teorema siguientes.

Lema 3.1 Sea una serie vectorial k -dimensional dada por $E_t = A_t + \Gamma(B)wI_t^{(h)}$, donde la

matriz de covarianza de A_t es $\Sigma = I$ y $w = m1_k$. Si S_E es la matriz de covarianza muestral de E_t normalizada por T , entonces,

$$E[S_E] = I + \frac{\beta_i}{T} m^2 1_k 1_k', \quad i = A, L, T$$

donde β_i está dado en cada caso por:

$$\beta_A = \frac{T-1}{T}, \quad \beta_L = \frac{(T-h+1)(h-1)}{T}, \quad \beta_T = \left(\frac{1-\delta^{2(T-h+1)}}{1-\delta^2} \right) - \frac{1}{T} \left(\frac{1-\delta^{(T-h+1)}}{1-\delta} \right)^2. \quad (3.2)$$

Este Lema prueba que de los tres atípicos considerados el cambio de nivel produce el mayor sesgo esperado en la matriz de covarianzas residual. Para estudiar el efecto de este sesgo en los estadísticos $J_{i,h}$ en (2.16), definimos los estadísticos muestrales $J_{i,h}^M$ obtenidos reemplazando en la expresión de $J_{i,h}$, Σ por $E[S_E]$. Se define la eficacia relativa del estadístico muestral mediante,

$$L_{i,k,h}(w) = \frac{E[J_{i,h}^M]}{E[J_{i,h}]}. \quad i = A, L, T \quad (3.3)$$

Por lo tanto, cuanto menor sea $L_{i,k,h}(w)$ menor será en términos relativos el estadístico muestral y mayor será la influencia del atípico en el valor del estadístico multivariante. A continuación, se obtiene el valor de $L_{i,k,h}(w)$ para $w = m1_k$, en el siguiente Teorema.

Teorema 3.2 Sea una serie vectorial k -dimensional dada por $E_t = A_t + \Gamma(B)wI_t^{(h)}$, donde la matriz de covarianza de A_t es $\Sigma = I$ y $w = m1_k$. Entonces, el valor de (3.3) para cada tipo de atípico está dado por:

1. Para los casos MIO y MAO,

$$L_{A,k,h}(m1_k) = 1 - \frac{\frac{\beta_A}{T} m^2}{(1+m^2)} \frac{(1+m^2 k)}{(1+m^2 \frac{\beta_A}{T} k)}.$$

2. Para el caso MLS,

$$L_{L,k,h}(m1_k) = 1 - \frac{\frac{\beta_L}{T} m^2}{(1+(T-h+1)m^2)} \frac{(1+(T-h+1)m^2 k)}{(1+m^2 \frac{\beta_L}{T} k)}.$$

3. Para el caso MTC,

$$L_{T,k,h}(m1_k) = 1 - \frac{\frac{\beta_T}{T} m^2}{\left(1 + \frac{(1+\delta)(1-\delta^{(T-h+1)})}{(1-\delta)(1+\delta^{(T-h+1)})} m^2 \right)} \frac{\left(1 + \frac{(1+\delta)(1-\delta^{(T-h+1)})}{(1-\delta)(1+\delta^{(T-h+1)})} m^2 k \right)}{(1+m^2 \frac{\beta_T}{T} k)}.$$

Tabla 3.1: Estimación de L cuando $h=T/2$.

	MAO					MLS					MTC				
$m \backslash k$	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	.995	.987	.979	.973	.965	.802	.667	.571	.499	.443	.990	.977	.965	.952	.941
2	.980	.957	.936	.915	.895	.502	.333	.250	.200	.166	.963	.926	.890	.859	.828
3	.957	.913	.875	.839	.806	.308	.181	.129	.099	.081	.921	.851	.791	.738	.691
4	.927	.860	.803	.752	.707	.199	.110	.076	.058	.047	.868	.764	.682	.616	.562
5	.890	.799	.725	.662	.611	.137	.073	.050	.038	.031	.809	.677	.582	.510	.453

En primer lugar, es fácil ver que $L_{i,k,h}(m1_k)$ es una función decreciente del número de componentes, k , suponiendo fijo el tamaño muestral, T . Esto significa que cuanto mayor es el número de componentes del vector de series, mayor es la influencia del atípico en el estadístico $J_{i,h}$. Si se supone que el número de componentes y el tamaño muestral crecen de forma que $k/T \rightarrow \nu$ y el momento de aparición del atípico y el tamaño muestral crecen de forma que $h/T \rightarrow \varphi$, entonces,

$$\begin{aligned}
L_{A,k,h}(m1_k) &\rightarrow 1 - \frac{m^4\nu}{(1+m^2)(1+m^2\nu)} \\
L_{L,k,h}(m1_k) &\rightarrow 1 - \frac{(1-\varphi)\varphi m^2\nu}{1+(1-\varphi)\varphi m^2\nu} \\
L_{T,k,h}(m1_k) &\rightarrow 1 - \frac{\frac{1+\delta}{1-\delta}m^4\nu}{\left(1+\frac{1+\delta}{1-\delta}m^2\right)(1+m^2\nu)}
\end{aligned}$$

y si además, $m \rightarrow \infty$, $L_{i,k,h}(m1_k) \rightarrow 0$. La Tabla 3.1 ilustra este resultado. Para cada tipo de atípico, se han generado 10000 series de ruido blanco con matriz de covarianzas I , para $k = 1, \dots, 5$ y tamaño muestral $T = 200$, contaminada por un atípico $w = m1_k$ para $m = 1, 2, 3, 4$ y 5 . Se calcula el estadístico $J_{i,h}$ para cada serie con $h = T/2$, primero, estimando la matriz de covarianzas y segundo, suponiendo que la matriz de covarianzas es conocida. Entonces, se calcula la media sobre 10000 replicas y se estima $L_{i,k,h}$ como sigue:

$$\widehat{L}_{i,k,h}(m1_k) = \frac{\overline{J_{i,h}^M}}{\overline{J_{i,h}}}, \quad i = A, L, T$$

La Tabla 3.1 muestra que cuanto mayor es la dimensión k , mayor es la pérdida de eficacia. El resultado es especialmente importante para cambios de nivel, como se esperaba del análisis previo.

3.3. Direcciones de proyección

Se ha demostrado que un atípico multivariante produce un atípico univariante en casi toda serie proyectada. Los algoritmos de Búsqueda de la Proyección (“Projection Pursuit”) tienen el objetivo de encontrar direcciones interesantes para la proyección de los datos. Estas proyecciones se obtienen mediante la maximización o minimización de una función objetivo denominada índice de proyección, que depende de los datos y del vector de proyección. El termino “proyección interesante” ha sido tradicionalmente asociado con proyecciones que muestran algún tipo de estructura inesperada, como conglomerados, atípicos o cualquier tipo de relación no lineal entre las variables. Es habitual suponer que las proyecciones más interesantes son las más lejanas de la normalidad. Algunas revisiones generales de las técnicas de Projection Pursuit se pueden encontrar en Huber (1985), Jones y Sibson (1987) y Posse (1995).

Peña y Prieto (2001, a) demuestran que dadas dos variables aleatorias vectoriales que tienen distribuciones simétricas con diferentes medias y la misma matriz de covarianzas, la dirección que minimiza el coeficiente de curtosis de la proyección es la función lineal discriminante, esto es, la dirección que produce la máxima separación entre las medias proyectadas con respecto a la varianza de la distribución de la proyección, y proponen un procedimiento para detección de atípicos multivariantes basado en proyecciones que maximizan y minimizan el coeficiente de curtosis de los datos proyectados. Peña y Prieto (2001, b) demuestran que estas direcciones son también útiles para identificar conglomerados en datos multivariantes.

Se define la dirección de máxima discriminación como aquella que maximiza el tamaño del atípico univariante, $v'w$, con respecto a la varianza de la serie proyectada. En esta sección se prueba que en los casos de atípico aditivo, cambio de nivel y cambio transitorio multivariantes, la dirección del atípico es la dirección de máxima discriminación y que esta dirección se obtiene encontrando los extremos del coeficiente de curtosis de la serie proyectada. En el caso de un atípico innovativo multivariante, se demuestra que la dirección del atípico es la de máxima discriminación para la serie de residuos, y que puede obtenerse buscando los extremos de la curtosis al proyectar los residuos.

Sea $Y_t = X_t + \alpha(B) w I_t^{(h)}$, $t = 1, \dots, T$, como en (2.12) y $E_t = A_t + \Gamma(B) w I_t^{(h)}$ como en (2.14). Se supone primero, para simplificar la presentación, y sin pérdida de generalidad porque el algoritmo propuesto en la siguiente sección es afín equivariante, que $E(X_t) = 0_{k \times 1}$ y $\Sigma_X = Cov(X_t) = I$. Se define la variable determinista,

$$R_t^{(h,T)} = \alpha(B) w I_t^{(h)} = w I_t^{(h)} - \alpha_1 w I_{t-1}^{(h)} - \dots - \alpha_{T-h} w I_{t-(T-h)}^{(h)}. \quad (3.4)$$

La variable $R_t^{(h,T)}$ tiene dos parámetros, el tiempo en el que el atípico aparece, $t = h$ y el tamaño muestral T y describe los efectos del atípico sobre la serie en todos los instantes posteriores al de ocurrencia. Proyectando Y_t en la dirección de v , se obtiene $y_t = x_t + r_t^{(h,T)}$, donde $r_t^{(h,T)} = v' R_t^{(h,T)}$. Se denota R_t y r_t a los coeficientes en tiempo t de las variables $R_t^{(h,T)}$

y $r_t^{(h,T)}$, respectivamente, esto es, a los efectos sobre la serie original y proyectada, dados por:

$$R_t = \begin{cases} 0 & t < h \\ -\alpha_{t-h}w & t \geq h \end{cases} \quad r_t = \begin{cases} 0 & t < h \\ -v'\alpha_{t-h}w & t \geq h \end{cases} \quad (3.5)$$

donde $\alpha_0 = -I$. También se denota, $\bar{R} = \frac{1}{T} \sum_{t=1}^T R_t$ y $\bar{r} = \frac{1}{T} \sum_{t=1}^T r_t$. Sean,

$$E \left[\frac{1}{T} \sum_{t=1}^T Y_t \right] = \frac{1}{T} (I - \alpha_1 - \dots - \alpha_{T-h}) w = \bar{R},$$

y,

$$\Sigma_Y = E \left[\frac{1}{T} \sum_{t=1}^T \left(Y_t - \frac{1}{T} \sum_{l=1}^T Y_l \right) \left(Y_t - \frac{1}{T} \sum_{l=1}^T Y_l \right)' \right] = I + \Sigma_R, \quad (3.6)$$

donde, $\Sigma_R = \frac{1}{T} \sum_{t=1}^T (R_t - \bar{R})(R_t - \bar{R})'$. Vamos a demostrar que la dirección del atípico es la de máxima discriminación. Utilizando resultados en Rao (1973, pg. 60), el máximo de $(v'w)^2 / (v'\Sigma_Y v)$ bajo la restricción $v'\Sigma_Y v = 1$ es $v = \Sigma_Y w$. En los casos de MAO, MLS y MTC, $\Sigma_Y = I + \beta_i w w'$, donde β_i están dados en (3.2), y entonces, $v = (1 + \beta_i w'w) w$, esto es, v es proporcional a w . Por el mismo resultado se generaliza al caso MIO para el máximo de $(v'w)^2 / (v'\Sigma_A v)$ bajo la restricción $v'\Sigma_A v = 1$. Esto demuestra que w es una dirección de máxima discriminación.

Suponiendo que v verifica $v'\Sigma_Y v = 1$, el coeficiente de curtosis de la serie y_t está dado por:

$$\gamma_y(v) = E \left[\frac{1}{T} \sum_{t=1}^T \left(y_t - \frac{1}{T} \sum_{l=1}^T y_l \right)^4 \right]. \quad (3.7)$$

Se demuestra ahora que w se puede obtener maximizando o minimizando el coeficiente de curtosis $\gamma_y(v)$, es decir, resolviendo los problemas de optimización:

$$\max_{s.t. \ v'\Sigma_Y v=1} \gamma_y(v) \quad \text{and} \quad \min_{s.t. \ v'\Sigma_Y v=1} \gamma_y(v). \quad (3.8)$$

Para encontrar las condiciones de primer orden para (3.8), se necesitan algunos resultados previos cuyas pruebas se pueden encontrar en el apéndice de este capítulo.

Lema 3.3 *El coeficiente de curtosis de y_t se puede escribir como sigue,*

$$\gamma_y(v) = 3(v'\Sigma_Y v)^2 - 3(v'\Sigma_R v)^2 + \omega_r(v), \quad (3.9)$$

donde $\omega_r(v) = \frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^4$.

Lema 3.4 *Las direcciones extremas del coeficiente de curtosis de y_t bajo la restricción $v' \Sigma_Y v = 1$ están dados por los autovectores de la matriz:*

$$\left[\sum_{t=1}^T \beta_t(v) B_t \right] v = \mu(v) v \quad (3.10)$$

donde $B_t = (R_t - \bar{R})(R_t - \bar{R})'$, y

$$\begin{aligned} \beta_t(v) &= (v' B_t v) - 3(v' \Sigma_R v) - \frac{\mu(v)}{T} \\ \mu(v) &= T(v' \Sigma_R v)^2 (\gamma_r(v) - 3), \end{aligned}$$

donde $\gamma_r(v)$ es el coeficiente de curtosis de $r_t^{(h,T)}$. Las direcciones que maximizan y minimizan el coeficiente de curtosis están dadas por los autovectores asociados al mayor y al menor de los autovalores $\mu(v)$, respectivamente.

El siguiente resultado demuestra la utilidad de las direcciones que maximizan y minimizan el coeficiente de curtosis de y_t .

Teorema 3.5 *Si X_t es un proceso ARMA(p, q) vectorial, e $Y_t = X_t + \alpha(B) w I_t^{(h)}$:*

1. *Para un MAO, el coeficiente de curtosis de y_t se maximiza cuando v es proporcional a w y se minimiza cuando v es ortogonal a w .*
2. *Para un MTC, el coeficiente de curtosis de y_t se maximiza o minimiza, respectivamente, cuando v es proporcional a w y se minimiza o maximiza, respectivamente, cuando v es ortogonal a w .*
3. *Para un MLS,*

- a) *el coeficiente de curtosis de y_t se minimiza cuando v es proporcional a w y se maximiza cuando v es ortogonal a w si:*

$$h \in \left(1 + \frac{1}{2} \left(1 - \frac{1}{\sqrt{3}} \right) T, 1 + \frac{1}{2} \left(1 + \frac{1}{\sqrt{3}} \right) T \right),$$

- b) *el coeficiente de curtosis de y_t se maximiza cuando v es proporcional a w y se minimiza cuando v es ortogonal a w si:*

$$h \notin \left(1 + \frac{1}{2} \left(1 - \frac{1}{\sqrt{3}} \right) T, 1 + \frac{1}{2} \left(1 + \frac{1}{\sqrt{3}} \right) T \right).$$

Por lo tanto, para un atípico aditivo multivariante, un cambio de nivel y un cambio transitorio, una de las direcciones que maximizan y minimizan el coeficiente de curtosis es asintóticamente la dirección del atípico. Observemos que estas direcciones se obtienen sin conocer el instante en que el atípico aparece. Como un atípico innovativo sólo afecta al residuo en tiempo $t = h$, es natural pensar que la dirección del atípico se obtendrá proyectando la serie de residuos. Esto se demuestra en el siguiente Corolario.

Corolario 3.6 *Sea X_t un proceso ARMA(p, q) vectorial estacionario, e, $Y_t = X_t + \Psi(B)wI_t^{(h)}$ y $E_t = A_t + wI_t^{(h)}$. El coeficiente de curtosis de $e_t = v'E_t$ se maximiza cuando v es proporcional a w y se minimiza cuando v es ortogonal a w .*

En la siguiente sección, se propone un algoritmo iterativo basado en las direcciones de máxima discriminación. Si están presentes atípicos de diferentes tipos, estas direcciones son complejas. Por ejemplo, para dos atípicos aditivos multivariantes, w_1 y w_2 , se puede demostrar que las direcciones extremas del coeficiente de curtosis de la serie vienen dadas, asintóticamente, por:

$$v = \frac{\beta_1^3}{\beta_1^4 + \beta_2^4} w_1 + \frac{\beta_2^3}{\beta_1^4 + \beta_2^4} w_2 \quad (3.11)$$

donde $\beta_1 = v'w_1$ y $\beta_2 = v'w_2$. Las direcciones w_1 y w_2 son soluciones de (3.11) y si w_1 y w_2 son ortogonales y suponiendo que $\|w_1\| > \|w_2\|$, entonces el máximo está dado en la dirección de w_1 . El mismo argumento se aplica al resto de los casos: las direcciones extremas del coeficiente de curtosis de la serie son asintóticamente una combinación lineal de las direcciones de los atípicos, incluyendo las direcciones de los atípicos como soluciones.

3.4. Algoritmo para la detección de atípicos

En esta sección, se propone un procedimiento secuencial para la detección de atípicos basado en las direcciones de máxima discriminación. El procedimiento se divide en cuatro pasos: (1) Obtener las direcciones de máxima discriminación y proyectar el vector de series en estas direcciones. (2) Buscar atípicos en las series univariantes obtenidas en dos etapas: primero, identificar cambios de nivel y segundo, identificar atípicos innovativos, aditivos y cambios transitorios. (3) Eliminar los efectos de los atípicos encontrados sobre la serie utilizando un modelo multivariante aproximado. (4) Repetir los pasos (1) a (3) sobre la nueva serie limpia de los efectos detectados en etapas anteriores, hasta que no se detecten más atípicos. Entonces, se identifica un modelo para la serie vectorial limpia se estiman conjuntamente los efectos y los parámetros en el modelo multivariante.

3.4.1. Cálculo de las direcciones de proyección

El primer paso en el algoritmo es el cálculo de las direcciones de proyección. Dada la serie temporal Y_t se necesitan resolver los problemas de optimización en (3.8). Se propone analizar un conjunto de $2k$ direcciones obtenidas maximizando y minimizando el coeficiente de curtosis de la serie proyectada directamente y sobre sus direcciones ortogonales. Para obtener estas direcciones, se considera el siguiente procedimiento, propuesto en Peña y Prieto (2001, b). Sea Y_t , la serie observada. Entonces:

1. Sean $i = 1$ y $Z_t^{(i)} = Y_t$.
2. Definir $\bar{Z}^{(i)} = \frac{1}{T} \sum_{t=1}^T Z_t^{(i)}$ y $\Sigma_Z^{(i)} = \frac{1}{T} \sum_{t=1}^T \left(Z_t^{(i)} - \bar{Z}^{(i)} \right) \left(Z_t^{(i)} - \bar{Z}^{(i)} \right)'$, y encontrar v_i tal que,

$$v_i = \arg \max_{v_i' \Sigma_Z^{(i)} v_i = 1} \frac{1}{T} \sum_{t=1}^T \left(v_i' Z_t^{(i)} - v_i' \bar{Z}^{(i)} \right)^4. \quad (3.12)$$

3. Si $i < k$, definir,

$$Z_t^{(i+1)} = \left(I - v_i v_i' \Sigma_Z^{(i)} \right) Z_t^{(i)},$$

esto es, $Z_t^{(i+1)}$ son las proyecciones de las observaciones en la dirección ortogonal a v_i . Sea $i = i + 1$. En otro caso, parar.

4. Repetir el mismo procedimiento minimizando la función objetivo en (3.12) para obtener otro conjunto de k direcciones, v_{k+1}, \dots, v_{2k} .

El principal punto en el algoritmo es resolver el problema en (3.12) y su correspondiente problema de minimización. Para ello, se utiliza un método de Newton modificado propuesto por Peña y Prieto (2001, b).

3.4.2. Búsqueda de atípicos univariantes

Como se mencionó en el capítulo 2, los contrastes más utilizados en la detección de atípicos en series temporales univariantes son los estadísticos de razón de verosimilitudes. Dada una serie temporal univariante y_t afectada por la presencia de un atípico en un punto temporal $t = h$, la serie filtrada de residuos está definida por:

$$e_t = a_t + \gamma(B) \beta I_t^{(h)}, \quad (3.13)$$

donde $\gamma(B) = 1 - \sum_{i=1}^{\infty} \gamma_i B^i$, y $\gamma(B) = 1$ para un atípico innovativo, $\gamma(B) = \pi(B)$ para un atípico aditivo, $\gamma(B) = (1 - B)^{-1} \pi(B)$ para un cambio de nivel y $\gamma(B) = (1 - \delta B)^{-1} \pi(B)$

para un cambio transitorio. El estadístico de razón de verosimilitudes para contrastar la hipótesis $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$ para cada tipo de atípico está dado por,

$$\lambda_{i,h} = \frac{\beta_{i,h}}{\rho_{i,h}\sigma_a}, \quad i = I, A, L, T \quad (3.14)$$

donde $\rho_{i,h}^2 = \left(\sum_{j=0}^{T-h} \gamma_j^2 \right)^{-1}$ con $\gamma_0 = -1$ y $\beta_{i,h} = -\rho_{i,h}^2 \left(\sum_{j=0}^{T-h} \gamma_j e_{h+j} \right)$ son las estimaciones de los tamaños de los atípicos. Como $\lambda_{i,h}^2$ son los estadísticos $J_{i,h}$ en el caso de $k = 1$, las distribuciones de $\lambda_{i,h}^2$ bajo la presencia de un atípico cuando los parámetros son conocidos son $\chi_1^2(\tilde{\eta}_i)$ donde $\tilde{\eta}_i = \left(\frac{\beta}{\rho_{i,h}\sigma_a} \right)^2$.

En la práctica, la localización del atípico y los parámetros del modelo son desconocidos. Entonces los parámetros son substituidos por sus estimaciones y se definen los siguientes estadísticos globales,

$$\Lambda(i, h_i) = \max_{1 \leq t \leq T} |\lambda_{i,t}|, \quad i = I, A, L, T.$$

Como comentamos en el capítulo 2, Chang y Tiao (1983) proponen un algoritmo iterativo basado en 3.14 para atípicos aditivos e innovativos, extendido a cambios de nivel y cambios transitorios por Tsay (1988).

Aquí se considera un enfoque diferente. Existe gran evidencia de que el contraste de razón de verosimilitudes confunde cambios de nivel con atípicos innovativos. Esto fue demostrado en Balke (1993) y más tarde analizado en Sánchez y Peña (2003), quienes demuestran que el valor crítico para el contraste de razón de verosimilitudes para cambios de nivel es diferente de los de los atípicos innovativos y aditivos. Por lo tanto, identificaremos primero cambios de nivel en las series antes de buscar el resto de atípicos. Para ello, es necesario un procedimiento capaz de detectar cambios de nivel robusto a los otros atípicos. Carnero, Peña y Ruiz (2003) demuestran que el contraste de razón de verosimilitudes para cambios de nivel no funciona bien para series temporales financieras y proponen el uso de un contraste de tipo cusum. Utilizando la notación introducida en la sección 3.2.1, Bai (1994) demuestra que el estadístico,

$$C_{h-1} = \frac{h-1}{\sqrt{T}\psi(1)\sigma_a} \left(\frac{1}{h-1} \sum_{t=1}^{h-1} y_t - \frac{1}{T} \sum_{t=1}^T y_t \right), \quad (3.15)$$

converge débilmente a un Puente Browniano en $[0, 1]$. Este estadístico, C_{h-1} es el utilizado para contrastar un cambio de nivel en $t = h$. En la práctica, el término $\psi(1)\sigma_a$ se reemplaza por un estimador consistente. Como $\psi(1)^2\sigma_a^2 = 2\pi f(0)$, donde $f(0)$ es la densidad espectral de la serie en la frecuencia 0, Bai (1994) recomienda el siguiente estimador,

$$\widehat{\psi(1)\sigma_a} = \left[\widehat{\gamma(0)} + 2 \sum_{i=1}^K \left(1 - \frac{|i|}{K} \right) \widehat{\gamma(i)} \right]^{\frac{1}{2}},$$

donde $\widehat{\gamma}(h)$ es la autocovarianza estimada de orden h y K es una cantidad tal que $K \rightarrow \infty$ y $K/T \rightarrow 0$ (ver Priestley, 1981). Supongamos que el cambio de nivel aparece en $t = h$, y sea $\kappa = h/T$ y $(\widehat{h} - 1) = \arg \max_{1 \leq t \leq T} |C_t|$. Entonces, $\widehat{\kappa} = \widehat{h}/T$ es el estimador del momento del cambio. Bai (1994) demuestra que $\widehat{\kappa} - \kappa = O(1/T)$, esto es, la consistencia del estimador del momento del cambio de nivel. El estadístico $\max_{1 \leq t \leq T} |C_t|$ bajo la hipótesis de que no hay cambios de nivel en la serie se distribuye asintóticamente como el supremo del valor absoluto de un Puente Browniano, W_t^0 , cuya función de distribución está dada por (ver, Billingsley, 1968),

$$F_{mC}(c) = 1 + 2 \sum_{i=1}^{\infty} (-1)^i e^{-2i^2 c^2}, \quad c > 0. \quad (3.16)$$

El estadístico cusum (3.15) tiene varias ventajas sobre el estadístico de razón de verosimilitudes para detectar cambios de nivel. Primero, la distribución asintótica es independiente de la distribución del error, por lo que no se requiere la hipótesis de Gaussianidad. Segundo, no es necesario especificar el orden del modelo ARMA. Tercero, como se verá en la sección 3.6, este estadístico parece ser más potente que el de razón de verosimilitudes en todos los modelos considerados. Cuarto, se demostrará que el estadístico (3.15) es robusto a la presencia de atípicos, mientras que el de razón de verosimilitudes no lo es.

3.4.2.1. Detección de cambios de nivel

Dadas las $2k$ series univariantes $y_{t,1} = v'_1 Y_t, \dots, y_{t,2k} = v'_{2k} Y_t$, se propone un procedimiento iterativo para identificar cambios de nivel basado en el algoritmo propuesto por Inclán y Tiao (1994) para cambios de varianza y Carnero, Peña y Ruiz (2003) para cambios de nivel en series temporales financieras de ruido blanco. El algoritmo está basado en dividir la serie en dos partes después de detectar un cambio, y procede como sigue:

1. Sea $t_1 = 1$.
2. Obtener,

$$D_L = \max_{1 \leq i \leq 2k} \max_{1 \leq t \leq T} |C_t^i|, \quad (3.17)$$

donde C_t^i está dado por (3.15) para $t = 1, \dots, T$ y las $2k$ series. Obtener,

$$(t_{\max}, i_{\max}) = \arg \max_{1 \leq i \leq 2k} \arg \max_{1 \leq t \leq T} |C_t^i|.$$

Si $|C_{t_{\max}}^{i_{\max}}| > D_{L,\alpha}$, asumir que existe un cambio de nivel para el vector de series en $t = t_{\max} + 1$, donde $D_{L,\alpha}$ es el valor crítico para un nivel de significatividad α e ir a 3.a. Si $|C_{t_{\max}}^{i_{\max}}| < D_{L,\alpha}$, suponer que no existe un cambio de nivel en la serie y terminar el algoritmo.

3.a Definir $t_2 = t_{\max}$ en el paso 2. Obtener,

$$(t_{\max}, i_{\max}) = \arg \max_{1 \leq i \leq 2k} \arg \max_{1 \leq t \leq t_2} |C_t^i|.$$

Si $|C_{t_{\max}}^{i_{\max}}| > D_{L,\alpha}$, se redefine $t_2 = t_{\max}$ y se repite el paso 3.a hasta que $|C_{t_{\max}}^{i_{\max}}| < D_{L,\alpha}$. Se define $t_{first} = t_2$ donde t_2 es el último valor que alcanza el máximo de los estadísticos cusum y es mayor que $D_{L,\alpha}$. El punto $t_{first} + 1$ es el primer punto con un posible cambio de nivel.

3.b Se repite una búsqueda similar en el intervalo $t_2 \leq t \leq T$, donde t_2 es el punto t_{\max} obtenido en el paso 2. Para ello, definir $t_1 = t_{\max} + 1$, donde,

$$(t_{\max}, i_{\max}) = \arg \max_{1 \leq i \leq 2k} \arg \max_{t_1 \leq t \leq T} |C_t^i|,$$

y repetir hasta que $|C_{t_{\max}}^{i_{\max}}| < D_{L,\alpha}$. Se define $t_{last} = t_1 - 1$, donde t_1 es el último valor que alcanza el máximo de los estadísticos cusum y es mayor que $D_{L,\alpha}$.

3.c Si $|t_{last} - t_{first}| < H$, donde H es un entero que define el menor intervalo entre dos cambios de nivel, existe únicamente un cambio y termina el algoritmo. Si no, mantener los dos valores como posibles cambios y repetir los pasos 2 y 3 para $t_1 = t_{first}$ y $T = t_{last}$, hasta que no se detecten más puntos de cambio. Entonces, ir al paso 4.

4. Definir un vector $\ell = (\ell_1, \dots, \ell_r)$ donde $\ell_1 = 1$, $\ell_r = T$ y $\ell_2, \dots, \ell_{r-1}$ sean los puntos detectados en los pasos 2 y 3 en orden creciente. Obtener el estadístico D_L en cada uno de los intervalos (ℓ_i, ℓ_{i+2}) y contrastar si el máximo de los estadísticos D_L es todavía significativo. Si no lo es, se elimina el correspondiente punto de posible cambio. Repetir el paso 4 hasta que el número de puntos de posible cambio no varíe y los puntos encontrados en el paso previo no difieran en H observaciones de donde estaban en la iteración previa. Se eliminan los puntos $\ell_1 = 1$ y $\ell_r = T$ del vector final. Se añade uno a los restantes valores, obteniendo los puntos con los cambios de nivel.

Un primer comentario sobre este algoritmo es que el estadístico $\max_{1 \leq t \leq T} |C_t|$ bajo la hipótesis de no existencia de cambios de nivel en la serie está distribuido asintóticamente como el supremo del valor absoluto de un Puente Browniano, W_t^0 , y éste tiene una función de distribución dada por (3.16). Se escribe,

$$D_L = \max_{1 \leq i \leq 2k} \max_{1 \leq t \leq T} |C_t^i| = \max_{j=1,2} \max_{1 \leq i_j \leq k} \max_{1 \leq t \leq T} |C_t^{i_j}|,$$

donde j es 1 para las k direcciones del máximo y es 2 para las k direcciones del mínimo. Entonces, D_L se puede escribir como el máximo de dos variables aleatorias dependientes y debido a esa dependencia, la distribución de D_L es intratable. En la sección siguiente, se obtendrán valores críticos mediante simulación para diferentes niveles de significatividad.

El segundo comentario es acerca del número H en los pasos 3.c y 4. Como los estadísticos (3.15) para observaciones cercanas están altamente correlados, podemos obtener varios máximos cuando existe un cambio de nivel. Por lo tanto, no se permite que dos cambios de nivel estén muy próximos. Se propone tomar $H = k + m$ donde m puede ser fijado por el usuario. En la implementación práctica, se ha utilizado $m = 10$.

El tercer comentario se refiere a la estimación de los cambios de nivel detectados. Sean $\{\ell_1, \dots, \ell_{r_L}\}$ los detectados. Para obtener una serie limpia de cambios de nivel, se ajusta el modelo:

$$(I - \Pi_1 B - \dots - \Pi_{\hat{p}} B^{\hat{p}}) Y_t^* = C + E_t^*, \quad (3.18)$$

donde $Y_t^* = Y_t - \sum_{i=1}^{r_L} w_i S_t^{(\ell_i)}$, y el orden \hat{p} se elige con que,

$$\hat{p} = \arg \min_{0 \leq p \leq p_{\max}} AIC(p) = \arg \min_{0 \leq p \leq p_{\max}} \left\{ \log |\hat{\Sigma}_p| + 2 \frac{k^2 p}{T} \right\},$$

donde $\hat{\Sigma}_p = \frac{1}{T-2p-1} \sum_{t=p+1}^T E_t^* E_t^{*'} y p_{\max}$ es una cota dada. Si alguno de los efectos no es significativo, se elimina de la ecuación (3.18) y se estiman los efectos de los restantes cambios de nivel. Este proceso se repite hasta que todos los cambios de nivel sean significativos.

3.4.2.2. Algoritmos para la detección del resto de atípicos

Después de obtener una serie libre del efecto de cambios de nivel, se propone el siguiente procedimiento de búsqueda de atípicos aditivos, innovativos y cambios transitorios:

1. Obtener las $2k$ direcciones que maximizan y minimizan el coeficiente de curtosis de la serie Y_t^* proyectada y sus direcciones ortogonales. Sean $y_{t,1}, \dots, y_{t,2k}$ las series obtenidas. Obtener otras $2k$ direcciones que maximizan y minimizan el coeficiente de curtosis de la serie proyectada E_t^* y sus direcciones ortogonales. Sean $e_{t,1}, \dots, e_{t,2k}$ las series obtenidas.
2. Para cada serie univariante $y_{t,1}, \dots, y_{t,2k}$, ajustar un modelo autorregresivo por medio del criterio de selección de Akaike (AIC). Para $t = 1, \dots, T$, calcular los estadísticos, $\lambda_{A,t}^i$ y $\lambda_{T,t}^i$, $i = 1, \dots, 2k$, utilizando las estimaciones de los parámetros de los ajustes autorregresivos. Obtener el máximo de los estadísticos $|\lambda_{A,t}^i|$ y $|\lambda_{T,t}^i|$ para cada serie, y a lo largo de la serie. Por otro lado, para cada serie residual univariante $e_{t,1}, \dots, e_{t,2k}$, calcular los estadísticos, $\lambda_{I,t}^i$, $i = 1, \dots, 2k$, obtener el máximo de los estadísticos $|\lambda_{I,t}^i|$ para cada serie, y a lo largo de la serie. Por lo tanto, se obtienen:

$$\Lambda_A = \max_{1 \leq i \leq 2k} \max_{1 \leq t \leq T} |\lambda_{A,t}^i| \quad \Lambda_T = \max_{1 \leq i \leq 2k} \max_{1 \leq t \leq T} |\lambda_{T,t}^i| \quad \Lambda_I = \max_{1 \leq i \leq 2k} \max_{1 \leq t \leq T} |\lambda_{I,t}^i|. \quad (3.19)$$

3. Sean $\Lambda_{A,\alpha}$, $\Lambda_{T,\alpha}$ y $\Lambda_{I,\alpha}$ los valores críticos para algún nivel de significatividad predeterminado α . Se tienen tres posibilidades:

- a) Si $\Lambda_j < \Lambda_{j,\alpha}$, $j = I, A, T$, no se detectan atípicos, e ir al paso 4.
- b) Si $\Lambda_j > \Lambda_{j,\alpha}$, para un sólo tipo $j = I, A, T$, se identifica un atípico del tipo j y se elimina su efecto usando las estimaciones de los parámetros multivariantes.
- c) Si $\Lambda_j > \Lambda_{j,\alpha}$, para más de un tipo, se identifica el atípico más significativo y se elimina su efecto utilizando las estimaciones de los parámetros multivariantes.

Se repite este paso, hasta que no se detectan más atípicos.

4. Sean $\{h_1^A, \dots, h_{r_A}^A\}$, $\{h_1^T, \dots, h_{r_T}^T\}$ y $\{h_1^I, \dots, h_{r_I}^I\}$ los tiempos de los r_A , r_T y r_I atípicos aditivos, innovativos y cambios transitorios detectados, respectivamente. Se estima conjuntamente los parámetros del modelo y los atípicos detectados para la serie Y_t^* :

$$(I - \Pi_1 B - \dots - \Pi_{\hat{p}} B^{\hat{p}}) Y_t^{**} = C + E_t^{**},$$

donde,

$$Y_t^{**} = Y_t^* - \sum_{i_A=1}^{r_A} w_{i_A} I_t^{(h_{i_A}^A)} - \sum_{i_T=1}^{r_T} \frac{w_{i_T}}{1 - \delta B} I_t^{(h_{i_T}^T)} \quad E_t^{**} = E_t^* - \sum_{i_I=1}^{r_I} w_{i_I} I_t^{(h_{i_I}^I)}.$$

Si alguno de los efectos no es significativo, se elimina el menos significativo. Este proceso se repite hasta que todos los atípicos en el conjunto final sea significativo.

Los valores críticos de los estadísticos $\lambda_{A,t}^i$, $\lambda_{T,t}^i$ y $\lambda_{I,t}^i$ deben ser obtenido mediante simulación. En la sección 3.6, se obtendrán varios valores críticos para diferentes modelos y número de componentes.

3.4.3. Estimación final conjunta

En las etapas anteriores se han obtenido un conjunto de cambios de nivel y atípicos significativos. Ahora, se procede a una estimación conjunta de los parámetros del modelo, los cambios de nivel y los atípicos. Esto es, se estima el modelo:

$$(I - \Phi_1 B - \dots - \Phi_{\hat{p}} B^{\hat{p}}) Z_t = C + (I - \Theta_1 B - \dots - \Theta_{\hat{q}} B^{\hat{q}}) B_t,$$

donde \hat{p} y \hat{q} son seleccionados a partir de la serie Y_t^{**} y,

$$Z_t = Y_t - \sum_{i_L=1}^{r_L} w_{i_L} S_t^{(i_L)} - \sum_{i_A=1}^{r_A} w_{i_A} I_t^{(h_{i_A}^A)} - \sum_{i_T=1}^{r_T} \frac{w_{i_T}}{1 - \delta B} I_t^{(h_{i_T}^T)} \quad B_t = E_t - \sum_{i_I=1}^{r_I} w_{i_I} I_t^{(h_{i_I}^I)},$$

y $\{\ell_1, \dots, \ell_{r_L}\}$, $\{h_1^A, \dots, h_{r_A}^A\}$, $\{h_1^I, \dots, h_{r_I}^I\}$ y $\{h_1^T, \dots, h_{r_T}^T\}$ son los tiempos de los r_L , r_A , r_I y r_T los cambios de nivel, atípicos aditivos, innovativos y cambios transitorios detectados, respectivamente. Si algún efecto (atípico o cambio de nivel) se encuentra significativo, se elimina el menos significativo y se repite la estimación conjunta hasta que todos los efectos sean significativos.

3.5. El caso no estacionario

Supongamos que $X_t \sim I(d_1, \dots, d_k)$, donde d_1, \dots, d_k son los grados de diferenciación en cada componente necesarios para convertirla en estacionaria. Se supone que $d_j \neq 0$ para al menos un j . Sea $d = \max(d_1, \dots, d_k)$ y supongamos en primer lugar que $d = 1$. Entonces consideramos los atípicos introducidos en Tsay, Peña y Pankratz (2000) y además un cambio de rampa multivariante (MRS) definido como sigue:

$$Y_t = X_t + wR_t^{(h)} \quad (3.20)$$

donde $R_t^{(h)} = (I - B)^{-1} S_t^{(h)}$. Este efecto es un cambio en la pendiente de la serie multivariante y puede aparecer en una serie $I(1)$. Un caso especial ocurre cuando $d_j = 1$ para algunas componentes mientras que las otras son $d_j = 0$. Los efectos w_j tendrán valor 0 en las componentes donde $d_j = 0$. Entonces, si alguna componente es $I(0)$ el cambio de rampa no le afecta.

El modelo para la serie X_t se transforma en estacionario tomando una diferencia. Se estudia a continuación su efecto sobre los atípicos. En el caso MIO, $(I - B)Y_t = (I - B)X_t + \tilde{\Psi}(B)wI_t^{(h)}$, donde $\tilde{\Psi}(B) = \nabla\Psi(B)$. Entonces, un MIO produce un MIO en la serie diferenciada. En el caso MAO, $(I - B)Y_t = (I - B)X_t + w(I_t^{(h)} - I_{t-1}^{(h)})$, produciendo dos MAO consecutivos con el mismo tamaño pero signo opuesto. En el caso del MLS, $(I - B)Y_t = (I - B)X_t + wI_t^{(h)}$, produciendo un MAO con el mismo tamaño. En el caso MTC, $(I - B)Y_t = (I - B)X_t + (I - B)(I - \delta B)^{-1}wI_t^{(h)} = (I - B)X_t + \zeta(B)wI_t^{(h)}$, donde $\zeta(B) = 1 + \zeta_1 B + \zeta_2 B^2 + \dots$ tal que $\zeta_j = \delta^{j-1}(1 - \delta)$. Entonces, un MTC produce un MTC con coeficientes de decrecimiento ζ_j . En el caso MRS, $(I - B)Y_t = (I - B)X_t + wS_t^{(h)}$ produciendo un MLS con el mismo tamaño.

Por lo tanto, en el caso $I(1)$, se propone un procedimiento similar al caso estacionario después de tomar una diferencia de la serie, como sigue:

1. Tomar una diferencia de la serie Y_t . A continuación, detectar MLS como en la sección 3.4.2.1. Los cambios de nivel detectados en la serie diferenciada se incorporan como cambios de rampa en la serie original y se estiman conjuntamente con los parámetros del modelo. Si alguno de los cambios de rampa no es significativo, se elimina y se reestima el modelo. Se repite este proceso hasta que todos los cambios de rampa sean significativos. Finalmente, se obtiene una serie $Y_t^* = Y_t - \sum_{i=1}^{r_R} w_i R_t^{(r_i)}$ sin cambios de rampa.
2. Se toma una diferencia de la serie Y_t^* y se obtiene el modelo:

$$(I - B)Y_t^* = (I - B)X_t + \eta(B)wI_t^{(h)}$$

donde $\eta(B) = \tilde{\Psi}(B)$ para un MIO, $\eta(B) = (I - B)$ para un MAO, $\eta(B) = I$ para un MLS y $\eta(B) = (I - B)(I - \delta B)^{-1}$ para un MTC. Entonces, se procede como en



la sección 3.4.2.2. Todos los atípicos detectados en la serie diferenciada se incorporan como los efectos correspondientes en la serie original y son estimados conjuntamente con los parámetros del modelo. Si alguno de los atípicos no es significativo, se elimina del modelo que es otra vez estimado. Se repite este proceso hasta que todos los atípicos presentes sean significativos.

Observemos que podría pensarse que el método no podría aplicarse en series cointegradas ya que el ajuste de autorregresivos de orden largo sería problemático al estar ∇Y_t sobrediferenciado y tener por tanto un componente de media móvil de raíz unidad. Sin embargo esto no es un problema para el método propuesto, ya que si v es un vector obtenido maximizando y minimizando el coeficiente de curtosis y no es un vector de cointegración, $v'\nabla Y_t = \nabla(v'Y_t)$ donde $v'Y_t$ es una serie no estacionaria y entonces, $\nabla(v'Y_t)$ es estacionaria e invertible, por lo que podemos buscar en ella los cambios de nivel y atípicos. Una vez que se hayan determinado estos cambios, se traslada la información de los atípicos a la serie original, esto es, un MLS en ∇Y_t es un MRS en Y_t , etc ... Entonces, en lugar del VAR sin restricciones, se ajusta un modelo autorregresivo de corrección del error para eliminar el problema de cointegración. El problema sólo aparece si v es el vector de cointegración, ya que entonces $v'Y_t$ es estacionario y $\nabla v'Y_t$ está sobrediferenciado. Sin embargo, hemos comprobado que no existe relación entre el vector que produce extremos en la curtosis proyectada y el vector de cointegración. Para ello, se han generado 10000 series de un modelo vectorial AR(1) de dos componentes con una relación de cointegración y se han buscado las direcciones en (3.12). Para compararlas con el vector de cointegración se ha calculado el valor absoluto del coseno del ángulo que forman. Los resultados demuestran que la media de los cosenos está alrededor de 0.62 con varianza 0.09, cuando si el ángulo fuese uniforme en el intervalo $(0, \pi)$, la distribución del coseno de dicho ángulo tiene esperanza 0.63 y varianza 0.09. A continuación, se repite el mismo experimento con las mismas series pero afectadas por atípicos, cambios de nivel o cambios transitorios y se obtiene que en cada caso la media de los ángulos tiende al ángulo que forma el atípico o su ortogonal, con la dirección de cointegración. Por lo tanto, se concluye que no parece existir ninguna relación entre los vectores de cointegración y las direcciones que maximizan y minimizan el coeficiente de curtosis.

Se considera ahora el caso $d = 2$, donde la serie es $I(2)$. Se define un cambio cuadrático multivariante como sigue:

$$Y_t = X_t + wQ_t^{(h)} \quad (3.21)$$

donde $Q_t^{(h)} = (I - B)^{-1} R_t^{(h)}$. Este efecto es un cambio en la tendencia cuadrática de la serie multivariante. La serie X_t puede ser transformada en estacionaria tomando dos diferencias. Entonces, un cambio cuadrático multivariante se transforma en un cambio de nivel, un cambio de nivel se transforma en un atípico aditivo multivariante, etc... Un procedimiento similar al propuesto para el caso $I(1)$ es aplicable a este caso. Este procedimiento se puede generalizar para cada orden de diferenciación d .

Tabla 3.2: Modelos para los experimentos.

$k = 2$			
Modelos	1	2	3
Φ	$\begin{pmatrix} 0.6 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}$	—	$\begin{pmatrix} 0.6 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}$
Θ	—	$\begin{pmatrix} -0.7 & 0 \\ -0.1 & -0.3 \end{pmatrix}$	$\begin{pmatrix} -0.7 & 0 \\ -0.1 & -0.3 \end{pmatrix}$
$k = 3$			
Modelos	4	5	6
Φ	$\begin{pmatrix} 0.6 & 0.2 & 0 \\ 0.2 & 0.4 & 0 \\ 0.6 & 0.2 & 0.5 \end{pmatrix}$	—	$\begin{pmatrix} 0.6 & 0.2 & 0 \\ 0.2 & 0.4 & 0 \\ 0.6 & 0.2 & 0.5 \end{pmatrix}$
Θ	—	$\begin{pmatrix} -0.7 & 0 & 0 \\ -0.1 & -0.3 & 0 \\ -0.7 & 0 & -0.5 \end{pmatrix}$	$\begin{pmatrix} -0.7 & 0 & 0 \\ -0.1 & -0.3 & 0 \\ -0.7 & 0 & -0.5 \end{pmatrix}$

3.6. Experimentos de Monte Carlo

En esta sección, se exploran los aspectos computacionales de los procedimientos propuestos. Primero, se obtienen valores críticos para todos los estadísticos considerados. Segundo, se comparan los estadísticos para detectar cambios de nivel. Por último, se presenta un estudio de simulación para estudiar la potencia de los estadísticos multivariantes y de proyección.

3.6.1. Valores críticos

Se consideran 6 modelos ARMA(p, q) vectoriales para los experimentos. El número de componentes es $k = 2$ y $k = 3$ y los parámetros matriciales aparecen en la Tabla 3.2. La constante en los modelos es siempre el vector 1_k . La matriz de varianzas residuales es la identidad.

Las dos matrices de parámetros autorregresivos tienen autovalores 0.27 y 0.72 para los modelos 1 y 3, y 0.27, 0.5 y 0.72, para los modelos 4 y 6, mientras que las matrices de parámetros de media móvil tienen autovalores -0.3 y -0.7 para los modelos 2 y 3, y -0.3 , -0.5 y -0.7 , para los modelos 5 y 6. A continuación, se generan valores críticos para los modelos 1 a 6. Los estadísticos considerados son:

$$\Lambda_j = \max_{1 \leq i \leq 2k} \max_{1 \leq t \leq T} |\lambda_{j,t}^i|, j = I, A, T \quad D_L = \max_{1 \leq i \leq 2k} \max_{1 \leq t \leq T} |C_t^i|. \quad (3.22)$$

Tabla 3.3: Valores críticos para los estadísticos considerados para $T=100, 200$ y 500 .

T	M	c.v.	Λ_I	Λ_A	Λ_L	Λ_T	D_L	c.v.	Λ_I	Λ_A	Λ_L	Λ_T	D_L
100	1	95 %	3.62	3.93	3.22	3.78	1.33	99 %	4.02	4.30	3.61	4.16	1.44
100	2	95 %	3.64	3.86	3.06	3.75	1.33	99 %	4.03	4.32	3.40	4.12	1.43
100	3	95 %	3.64	3.65	3.27	3.76	1.33	99 %	3.96	4.08	3.71	4.14	1.44
100	4	95 %	3.93	4.10	3.34	3.92	1.36	99 %	4.27	4.51	3.81	4.33	1.46
100	5	95 %	3.97	4.20	3.23	3.92	1.36	99 %	4.32	4.62	3.72	4.35	1.46
100	6	95 %	3.87	3.89	3.36	3.98	1.36	99 %	4.29	4.32	3.82	4.46	1.46
200	1	95 %	3.81	3.98	3.30	3.89	1.40	99 %	4.13	4.39	3.72	4.23	1.53
200	2	95 %	3.82	3.95	3.10	3.93	1.40	99 %	4.23	4.40	3.59	4.33	1.51
200	3	95 %	3.79	3.84	3.34	3.87	1.40	99 %	4.08	4.14	3.71	4.20	1.52
200	4	95 %	4.10	4.22	3.38	4.06	1.40	99 %	4.39	4.68	3.78	4.72	1.56
200	5	95 %	4.11	4.33	3.20	4.06	1.42	99 %	4.49	4.81	3.70	4.60	1.57
200	6	95 %	4.14	4.00	3.41	4.04	1.42	99 %	4.54	4.34	3.79	4.68	1.56
500	1	95 %	4.08	4.18	3.41	4.19	1.44	99 %	4.52	4.65	3.91	4.64	1.61
500	2	95 %	4.14	4.17	3.21	4.15	1.43	99 %	4.55	4.62	3.77	4.50	1.59
500	3	95 %	4.06	4.00	3.43	4.17	1.43	99 %	4.49	4.40	3.86	4.52	1.59
500	4	95 %	4.32	4.39	3.48	4.39	1.46	99 %	4.75	4.80	3.93	4.87	1.63
500	5	95 %	4.26	4.42	3.38	4.33	1.47	99 %	4.79	4.87	3.83	4.76	1.66
500	6	95 %	4.28	4.22	3.58	4.38	1.49	99 %	4.68	4.58	3.99	4.70	1.63

Se incluyen los estadísticos de proyección para el cambio de nivel para compararlo con el estadístico cusum D_L . Para calcular los estadísticos Λ_j y D_L , se procede como sigue. Para cada modelo y tamaño muestral $T = 100, 200$ y 500 , se han generado 10000 series. Se obtienen el conjunto de $2k$ direcciones ortogonales maximizando y minimizando el coeficiente de curtosis de la serie proyectada Y_t . Para $t = 1, \dots, T$, se calculan los estadísticos C_t^i , $i = 1, \dots, 2k$ y obtener D_L en (3.22). Para los otros estadísticos, para cada una de las $2k$ series, se ajusta un modelo autorregresivo mediante el criterio de información de Akaike (AIC). Se calculan los estadísticos, $\lambda_{A,t}^i$, $\lambda_{L,t}^i$ y $\lambda_{T,t}^i$ para $t = 1, \dots, T$, $i = 1, \dots, 2k$ y se obtienen Λ_A , Λ_L and Λ_T en (3.22). Finalmente, para el atípico innovativo multivariante, para las 10000 series generadas previamente, se ajusta un modelo autorregresivo por medio del criterio de información de Akaike y se obtienen el conjunto de $2k$ direcciones ortogonales maximizando y minimizando el coeficiente de curtosis de los residuos proyectados, E_t . Para $t = 1, \dots, T$, se calculan los estadísticos, $\lambda_{I,t}^i$ y se obtienen Λ_I . Los valores críticos aparecen en la Tabla 3.3. Se observa que para T fijo los valores críticos de Λ_I , Λ_A y Λ_T son similares y mayores que los de Λ_L . Los valores críticos aumentan con T , como sería esperable.

3.6.2. Comparación de los estadísticos multivariantes, los estadísticos de proyección y el estadístico cusum para cambio de nivel

A continuación, se comparan los estadísticos multivariantes, los estadísticos de proyección y el estadístico cusum para cambio de nivel. Con este objetivo se han generado 1000 series con tamaños muestrales $T = 100$ y 200 contaminadas por un cambio de nivel en $h_L = 50$ y 100 respectivamente con tamaños $w_L = 3 \times 1_k$, $4 \times 1_k$, y direcciones aleatorias para cada modelo. Cada dirección aleatoria se obtiene mediante la generación de un valor de la distribución uniforme, u , para cada componente y definiendo $w_{Li} = -1, 0$ o 1 dependiendo de si u pertenece a los intervalos $(0, 1/3)$, $(1/3, 2/3)$ o $(2/3, 1)$. Después de esto, el vector w_L es normalizado para tener la misma norma que el vector $3 \times 1_k$. Se obtienen las proyecciones que maximizan y minimizan el coeficiente de curtosis de la serie proyectada y sus direcciones ortogonales. Entonces, se calcula el máximo de los estadísticos multivariantes (2.18) para un cambio de nivel, el máximo del estadístico de proyección (3.22) para un cambio de nivel y el máximo de los estadísticos cusum (3.22). Se comparan con sus respectivos valores críticos al nivel de significatividad del 95 %. La proporción de veces donde los estadísticos son mayores que su valor crítico se muestra en la primera y segunda parte de la Tabla 3.4 (ver columnas J_{\max} , Λ_L y D_L). En todos los modelos considerados, el contraste cusum es más potente que los otros estadísticos.

Ahora, se estudia la potencia de estos tres estadísticos bajo la presencia de cambios de nivel y de otros atípicos. Para ello, se han generado 1000 series con tamaño $T = 100$, contaminadas con un atípico innovativo en $h_I = 20$ de tamaño $w_I = w \times 1_k$, un atípico aditivo en $h_A = 40$ con tamaño $w_A = -w \times 1_k$, un cambio transitorio en $h_T = 80$ con tamaño $w_T = -w \times 1_k$ y un cambio de nivel en $h_L = 60$ con tamaño $w_L = w \times 1_k$, donde $w = 3$ y 4 , o una dirección aleatoria con la misma norma que $3 \times 1_k$, para cada modelo y se calcula el máximo de los estadísticos. Se comparan los estadísticos para cambio de nivel con sus respectivos valores críticos al 95 % de significatividad. La proporción de veces donde los estadísticos son mayores que sus valores críticos aparecen en la tercera parte de la Tabla 3.4 (ver columnas J_{\max} , Λ_L y D_L). Como en el caso de sólo cambios de nivel, la potencia del estadístico cusum es mayor que la potencia de otros estadísticos en todos los casos. Se ha medido la pérdida de potencia de cada contraste debido a los atípicos por,

$$\text{pérdida}(i) = 1 - \frac{\text{potencia con atípicos en el modelo } i}{\text{potencia sin atípicos en el modelo } i},$$

y se obtienen la media de la pérdida para los 6 modelos para cada estadístico para $w = 3$: la pérdida media para el estadístico multivariante es de un 27.7 %, la pérdida media para la proyección es 17.1 %, mientras que para el cusum es 9.4 %. Por lo tanto, los estadísticos multivariantes y de proyección están mucho más afectados por atípicos que el cusum.

Por último, es importante estudiar el error de tipo I en los estadísticos bajo la presencia de los otros atípicos. Para ello, se han generado 1000 series con tamaño $T = 100$, contaminadas

con un atípico innovativo en $h_I = 25$ con tamaño $w_I = w \times 1_k$, un atípico aditivo en $h_A = 50$ con tamaño $w_A = -w \times 1_k$, y un cambio transitorio en $h_T = 75$ con tamaño $w_T = w \times 1_k$, donde $w = 3, 4$ y direcciones aleatorias con la misma norma que $3 \times 1_k$ para cada modelo, se calculan los estadísticos multivariante, de proyección y cusum y se comparan con sus respectivos valores críticos al 95 % de significatividad. Las seis últimas filas en la Tabla 4.4 dan la frecuencia en que los estadísticos son mayores que los correspondientes valores críticos. La frecuencia de detección de un falso cambio de nivel para los estadísticos multivariantes para el modelo 6 y $w = 3$ y 4 son 16.6 % y 16.3 % respectivamente, para el estadístico de proyección para el modelo 3 y $w = 5$ es el 57 %, mientras que la frecuencia de detección para el estadístico cusum es siempre menor que el valor nominal del 5 %. Se concluye que los estadísticos multivariante y de proyección parecen no ser robustos a la presencia de otros atípicos.

3.6.3. Comparación de la potencia de los estadísticos multivariante y de proyección para detectar atípicos

En esta sección, se investiga la potencia de los estadísticos para detectar atípicos. Se han generado 1000 series con $T = 100$ y 200, para todos los modelos considerados afectados por la presencia de un atípico aditivo en $h_A = 50$ y 100, respectivamente, otros conjuntos de 1000 series con $T = 100$ y 200 afectados por la presencia de un atípico innovativo en $h_I = 50$ y 100, respectivamente, y otros conjuntos de 1000 series con $T = 100$ y 200 afectados por la presencia de un cambio transitorio en $h_T = 50$ y 100, respectivamente. Para cada serie y caso, se obtienen las proyecciones que maximizan y minimizan el coeficiente de curtosis de la serie proyectada y sus direcciones ortogonales. Entonces, se calculan los estadísticos multivariantes en (2.16) y el máximo de los estadísticos de proyección para el caso correspondiente en (3.22). A continuación, se compara cada valor con su respectivo valor crítico con el 95 % de significatividad. La Tabla 3.5 muestra las frecuencias de detección de los atípicos para los estadísticos multivariante y de proyección. En casi todos los casos, el estadístico de proyección es más potente que el multivariante.

3.7. Ejemplo

En esta sección, se ilustra el comportamiento del procedimiento propuesto mediante una serie de datos reales. Los datos corresponden a la tasa de crecimiento del producto nacional bruto de Francia, Italia y España durante el siglo XX. Los datos se pueden obtener de la página web <http://www.fgn.unisg.ch/eurmacro/>. El gráfico de las series aparece en la Figura 3.2.

Primero, se calculan las direcciones de proyección utilizando el procedimiento en la sección 3.4 y se aplica el algoritmo de detección de cambios de nivel. En este caso, el valor

Tabla 3.4: Frecuencia de detección de un cambio de nivel para los estadísticos multivariante (J_{\max}), de proyección (Λ_L) y cusum (D_L).

T	M	$w = 3 \times 1_k$							$w = 4 \times 1_k$			$w = aleat.$		
		h_I	h_A	h_L	h_T	J_{\max}	Λ_L	D_L	J_{\max}	Λ_L	D_L	J_{\max}	Λ_L	D_L
100	1	—	—	50	—	70.0	83.0	100	96.6	98.2	100	68.9	83.9	100
100	2	—	—	50	—	58.3	82.6	100	89.2	96.9	100	67.6	94.0	100
100	3	—	—	50	—	46.5	73.3	87.6	91.5	93.2	100	66.9	81.6	95.3
100	4	—	—	50	—	93.6	92.7	99.6	100	98.8	100	83.6	87.6	99.7
100	5	—	—	50	—	68.1	98.8	100	94.7	99.7	100	76.9	91.3	100
100	6	—	—	50	—	86.3	86.6	88.7	98.9	99.0	99.2	78.2	77.2	96.7
200	1	—	—	100	—	80.7	95.4	100	98.1	99.6	100	90.3	91.0	100
200	2	—	—	100	—	92.2	95.6	100	97.0	99.3	100	88.3	98.0	100
200	3	—	—	100	—	78.1	90.6	99.5	97.1	98.4	100	80.9	84.6	100
200	4	—	—	100	—	98.7	98.9	100	100	100	100	90.6	94.0	100
200	5	—	—	100	—	85.8	99.8	100	99.0	100	100	89.3	98.7	100
200	6	—	—	100	—	97.5	97.8	100	99.6	99.7	100	90.3	83.6	100
100	1	20	40	60	80	45.6	76.0	92.6	73.0	91.8	99.3	55.2	70.9	95.7
100	2	20	40	60	80	52.0	87.6	100	74.3	96.7	100	58.9	77.6	99.7
100	3	20	40	60	80	17.0	53.6	66.6	40.3	84.3	86.6	44.1	49.2	80.9
100	4	20	40	60	80	73.6	74.0	90.3	92.0	85.0	100	67.9	73.9	94.0
100	5	20	40	60	80	61.3	81.6	100	71.3	92.0	100	65.9	82.9	99.7
100	6	20	40	60	80	63.6	55.3	74.3	87.0	80.0	95.0	58.2	49.5	85.3
100	1	25	50	—	75	4.0	14.0	3.3	2.6	26.3	3.3	3.0	21.0	5.0
100	2	25	50	—	75	0.3	9.0	3.0	1.3	11.0	2.0	1.3	17.4	5.3
100	3	25	50	—	75	4.0	21.3	4.3	2.3	31.3	4.6	9.0	41.4	6.3
100	4	25	50	—	75	6.3	15.6	3.0	7.0	28.0	6.3	10.4	24.4	6.0
100	5	25	50	—	75	3.3	8.0	4.0	2.3	12.6	4.3	4.3	15.4	5.0
100	6	25	50	—	75	16.3	22.3	4.3	16.6	27.3	4.6	15.4	37.8	4.7

Tabla 3.5: Frecuencia de detección para los estadísticos multivariante y de proyección.

		$w = 3 \times 1_k$			$w = 4 \times 1_k$			$w = aleat.$	
	T	M	h_I	J_{\max}	Λ_I	J_{\max}	Λ_I	J_{\max}	Λ_I
MIO	100	1	50	59.9	77.8	95.0	98.0	63.2	80.6
	100	2	50	53.9	71.5	89.8	95.3	60.9	77.6
	100	3	50	51.1	68.6	88.0	95.2	47.5	76.6
	100	4	50	81.9	91.5	99.6	99.2	82.9	94.6
	100	5	50	62.0	76.1	95.1	97.4	73.6	86.6
	100	6	50	61.8	78.1	92.3	97.6	71.6	89.3
MIO	200	1	100	58.8	71.8	92.6	96.8	59.5	74.9
	200	2	100	58.5	68.4	92.5	95.5	56.9	72.9
	200	3	100	57.0	68.3	92.7	95.6	55.5	75.3
	200	4	100	81.6	87.5	100	100	80.6	91.3
	200	5	100	67.4	76.2	97.7	98.5	77.3	86.0
	200	6	100	67.1	75.8	98.2	98.8	78.6	86.0
	T	M	h_A	J_{\max}	Λ_A	J_{\max}	Λ_A	J_{\max}	Λ_A
MAO	100	1	50	86.6	93.6	99.3	99.3	80.6	84.6
	100	2	50	67.0	96.0	96.0	100	63.6	84.3
	100	3	50	91.0	99.0	99.6	100	87.3	95.7
	100	4	50	98.3	99.3	99.6	100	91.6	94.0
	100	5	50	78.6	95.6	98.6	100	78.6	92.3
	100	6	50	97.0	99.0	99.6	100	92.0	92.3
MAO	200	1	100	87.3	93.6	99.3	100	71.9	75.6
	200	2	100	67.0	91.0	95.6	99.3	64.6	83.3
	200	3	100	98.0	98.0	100	100	92.6	95.7
	200	4	100	99.3	99.6	100	100	97.3	93.6
	200	5	100	82.6	89.6	99.0	99.0	85.0	89.6
	200	6	100	98.8	97.8	100	100	97.0	95.6
	T	M	h_T	J_{\max}	Λ_T	J_{\max}	Λ_T	J_{\max}	Λ_T
MTC	100	1	50	61.3	88.6	93.3	98.6	60.9	97.7
	100	2	50	64.5	97.0	94.0	100	59.5	97.0
	100	3	50	71.3	93.6	92.6	99.3	57.2	95.7
	100	4	50	90.0	98.0	100	100	86.0	98.0
	100	5	50	71.0	97.6	95.3	99.6	79.9	99.3
	100	6	50	82.6	95.6	97.0	98.6	76.9	97.3
MTC	200	1	100	61.0	88.6	98.0	98.0	68.6	89.3
	200	2	100	66.0	92.0	94.6	99.3	68.9	92.0
	200	3	100	73.6	90.3	97.7	98.6	75.9	84.0
	200	4	100	92.0	92.8	99.5	99.5	92.6	94.7
	200	5	100	75.3	93.3	99.3	100	88.0	96.7
	200	6	100	95.9	95.7	100	99.2	85.0	92.6

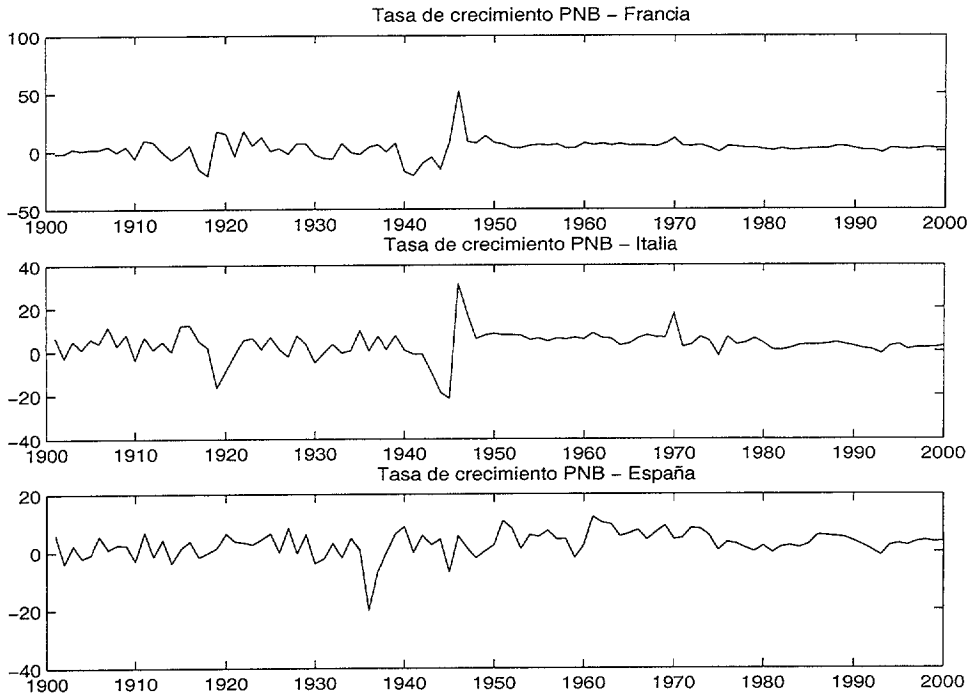


Figura 3.2: Tasa de crecimiento del producto nacional bruto de Francia, Italia y España.

crítico utilizado es 1.36. El algoritmo detecta un cambio en el instante $\ell_1 = 1946$. El valor del estadístico (3.17) para dicho tiempo es 1.57. Se estima su efecto ajustando un modelo autorregresivo mediante el AIC. Se elimina dicho efecto y se pasa a detectar otros atípicos. Los valores críticos utilizados son 3.92, 4.06 y 3.94 para los atípicos innovativo, aditivo y cambio transitorio respectivamente. Los atípicos encontrados por el algoritmo aparecen en la Tabla 3.6. Son detectados ocho atípicos por el procedimiento. El modelo seleccionado por el AIC es un modelo AR(1) vectorial, dado por:

$$X_t = \begin{pmatrix} 3.29 \\ 2.29 \\ 1.98 \end{pmatrix} + \begin{pmatrix} 0.44 & -0.54 & 0 \\ 0.32 & 0 & 0 \\ 0.13 & -0.15 & 0.31 \end{pmatrix} X_{t-1} + A_t \quad \hat{\Sigma} = \begin{pmatrix} 56.55 & 18.71 & 2.99 \\ 18.71 & 33.50 & 6.15 \\ 2.99 & 6.15 & 16.00 \end{pmatrix}.$$

Los tamaños de los ocho atípicos estimados están dados en la Tabla 3.7 junto con los t-ratios de los tamaños estimados. Todos los atípicos detectados están relacionados con hechos bien conocidos. Los de 1917 y 1919 aparecen durante y al final de la primera guerra mundial. Ambos atípicos afectan principalmente a Francia e Italia debido a que España no participó en aquel conflicto. El atípico en 1936 coincide con el comienzo de la guerra civil española y sólo afecta a la componente española. El resto de los atípicos aparecen en la segunda guerra mundial. Se hace notar que en los años 1945 y 1946, el procedimiento detecta dos atípicos en el mismo punto. Esto puede ser explicado porque un efecto externo puede afectar de dos

Tabla 3.6: Atípicos detectados por el algoritmo de proyección.

Iteraciones	(Λ_I, h_I)	(Λ_A, h_A)	(Λ_T, h_T)	Atípico	
				Año	Tipo
1	(5.56,1946)	(9.00,1946)	(8.12,1946)	1946	MAO
2	(5.37,1936)	(5.90,1936)	(5.94,1919)	1919	MTC
3	(5.32,1936)	(6.29,1945)	(5.24,1936)	1945	MAO
4	(5.41,1936)	(6.34,1936)	(5.61,1936)	1936	MAO
5	(3.68,1944)	(3.92,1944)	(4.24,1940)	1940	MTC
6	(4.04,1944)	(4.35,1918)	(5.03,1917)	1917	MIO
7	(4.56, 1946)	(4.81,1946)	(5.03,1945)	1945	MTC
8	(3.95,1944)	(4.03,1944)	(3.71,1961)	1944	MAO
9	(3.22,1915)	(3.83,1970)	(3.73,1961)	-	-

Tabla 3.7: Estimación de los tamaños de los atípicos detectados por el algoritmo de proyección.

Año	Tipo	\hat{w}_1 (<i>t-ratio</i>)	\hat{w}_2 (<i>t-ratio</i>)	\hat{w}_3 (<i>t-ratio</i>)	Año	Tipo	\hat{w}_1 (<i>t-ratio</i>)	\hat{w}_2 (<i>t-ratio</i>)	\hat{w}_3 (<i>t-ratio</i>)
1917	MIO	-20.5650 (-4.2965)	-2.6168 (-0.6515)	-3.2756 (-1.1527)	1945	MAO	3.5192 (0.6710)	-13.6251 (-3.2224)	-7.5548 (-2.0075)
1919	MTC	21.6840 (3.9502)	-12.5705 (-2.8224)	3.9320 (1.1230)	1945	MTC	14.4070 (3.5571)	1.4953 (0.3686)	-2.3170 (-0.8207)
1936	MAO	4.0151 (0.7435)	-5.5785 (-1.3803)	-17.8155 (-4.8810)	1946	MAO	29.0448 (4.3504)	22.1332 (4.3807)	2.0547 (0.5499)
1940	MTC	-21.4018 (-3.9794)	-5.6718 (-1.4000)	2.7293 (0.9516)	1946	MLS	2.118 (2.305)	1.744 (2.065)	1.086 (1.742)
1944	MAO	-8.9874 (-2.1508)	-16.6422 (-3.8230)	-0.1832 (-0.0589)					

maneras diferentes a componentes diferentes. Por ejemplo, en 1945, la serie para Francia está afectada por un MAO pero la serie para Italia está afectada por un MTC.

Finalmente, se comparan estos resultados con los obtenidos aplicando el procedimiento en Tsay, Peña y Pankratz (2000) que se resume en la Tabla 3.8. El procedimiento identifica cuatro atípicos innovacionales y cuatro cambios transitorios pero no identifica el cambio de nivel detectado por el procedimiento de las proyecciones en $t = 1946$. En este año, el procedimiento identifica un cambio transitorio multivariante y un atípico innovacional multivariante. Como Balke (1993) apuntó, un procedimiento que compara los estadísticos para cambio de nivel con los estadísticos para los otros atípicos debe confundir cambios de nivel con atípicos innovativos y esto es lo que puede ocurrir en este ejemplo.

Tabla 3.8: Atípicos detectados por el procedimiento de Tsay, Peña y Pankratz (2000).

Iteraciones	(J_I, h_I)	(J_A, h_A)	(J_L, h_L)	(J_T, h_T)	Atípico	
					Año	Tipo
1	(31.88,1946)	(28.75,1946)	(6.18,1946)	(31.51,1919)	1919	MTC
2	(28.34,1936)	(23.70,1936)	(5.84,1946)	(23.22,1946)	1936	MIO
3	(25.56,1946)	(23.35,1646)	(6.22,1946)	(24.94,1946)	1946	MTC
4	(20.87,1944)	(15.46,1945)	(6.57,1951)	(23.79,1940)	1940	MTC
5	(22.92,1944)	(15.95,1970)	(6.06,1951)	(15.54,1943)	1944	MIO
6	(22.14,1945)	(16.09,1945)	(6.07,1951)	(15.93,1917)	1945	MIO
7	(18.28,1918)	(17.44,1918)	(5.83,1951)	(20.07,1917)	1917	MTC
8	(17.77,1946)	(15.50,1970)	(5.41,1951)	(14.80,1915)	1946	MIO
9	(15.76,1970)	(16.99,1970)	(5.22,1951)	(14.79,1915)	-	-

3.8. Conclusiones

En este capítulo, se han analizado atípicos en series temporales multivariantes. Para ello, se ha adoptado un enfoque basado en proyecciones. Se ha demostrado la utilidad de obtener combinaciones lineales de un vector de series afectado por atípicos y que estadísticos univariantes aplicados a las series proyectadas pueden ser más potentes que los correspondientes estadísticos multivariantes. A continuación, dada una serie multivariante afectada por un atípico, se han analizado las direcciones que maximizan y minimizan el coeficiente de curtos de la serie proyectada y se ha demostrado que alguna de estas direcciones corresponde a la dirección del atípico. Basado en ello, se ha propuesto un algoritmo de detección y estimación de atípicos en series multivariantes estructurado en varios pasos. En primer lugar, se detectan cambios de nivel mediante un estadístico cusum para a continuación detectar el resto de atípicos mediante estadísticos de razón de verosimilitudes. Un aspecto de gran importancia y que demuestra la utilidad del método propuesto es que es innecesario construir un modelo multivariante completo para detectar los atípicos, algo inusual en este tipo de métodos. Un estudio de simulación demuestra la eficacia del método propuesto. Por último el método se ilustra mediante la serie de la tasa de crecimiento del producto nacional bruto de Francia, Italia y España durante el siglo veinte.

3.9. Apéndice

Demostración del Lema 3.1. La matriz de covarianzas muestral de los residuos, S_E , está dada por:

$$\begin{aligned} S_E = S_A + \frac{m}{T} [1_k A'_h - \Gamma_1 1_k A'_{h+1} - \dots - \Gamma_{T-h} 1_k A'_T] + \\ + \frac{m}{T} [A_h 1'_k - A_{h+1} 1'_k \Gamma'_1 - \dots - A_T 1'_k \Gamma'_{T-h}] + \frac{m^2}{T} [1_k 1'_k + \Gamma_1 1_k 1'_k \Gamma'_1 + \dots + \Gamma_{T-h} 1_k 1'_k \Gamma'_{T-h}] - \\ - \frac{m^2}{T^2} [I - \Gamma_1 - \dots - \Gamma_{T-h}] 1_k 1'_k [I - \Gamma_1 - \dots - \Gamma_{T-h}]'. \end{aligned}$$

Tomando esperanzas, y substituyendo la matriz polinomial $\Gamma(B)$ en cada caso, se obtienen las expresiones en (3.2). ■

Demostración del Teorema 3.2. Partiendo de,

$$E[S_E]^{-1} = \left(I + m^2 \frac{\beta_i}{T} 1_k 1'_k \right)^{-1} = I - \frac{m^2 \frac{\beta_i}{T}}{1 + m^2 \frac{\beta_i}{T} k} 1_k 1'_k, \quad i = I, A, L, T.$$

1. Para los casos MIO y MAO, $J_{A,h}^A = E'_h E[S_E]^{-1} E_h$, que resulta ser,

$$\begin{aligned} L_{A,k,h}(m 1_k) &= \frac{E[E'_h E[S_E]^{-1} E_h]}{E[E'_h E_h]} = 1 - \frac{m^2 \frac{\beta_A}{T}}{1 + m^2 \frac{\beta_A}{T} k} \frac{E[E'_h 1_k 1'_k E_h]}{E[E'_h E_h]} = \\ &= 1 - \frac{\frac{\beta_A}{T} m^2}{(1 + m^2)} \frac{(1 + m^2 k)}{(1 + m^2 \frac{\beta_A}{T} k)}. \end{aligned}$$

2. Para el caso MLS, $J_{L,h}^A = \frac{1}{T-h+1} \left(\sum_{t=h}^T E_t \right)' E[S_E]^{-1} \left(\sum_{t=h}^T E_t \right)$, que conduce a,

$$\begin{aligned} L_{L,k,h}(m 1_k) &= \frac{E \left[\left(\sum_{t=h}^T E_t \right)' E[S_E]^{-1} \left(\sum_{t=h}^T E_t \right) \right]}{E \left[\left(\sum_{t=h}^T E_t \right)' \left(\sum_{t=h}^T E_t \right) \right]} = \\ &= 1 - \frac{m^2 \frac{\beta_L}{T}}{1 + m^2 \frac{\beta_L}{T} k} \frac{E \left[\left(\sum_{t=h}^T E_t \right)' 1_k 1'_k \left(\sum_{t=h}^T E_t \right) \right]}{E \left[\left(\sum_{t=h}^T E_t \right)' \left(\sum_{t=h}^T E_t \right) \right]} = \\ &= 1 - \frac{m^2 \frac{\beta_L}{T}}{(1 + (T-h+1)m^2)} \frac{(1 + (T-h+1)m^2 k)}{(1 + m^2 \frac{\beta_L}{T} k)}. \end{aligned}$$

3. Para el caso MTC, $J_{T,h}^A = \frac{1-\delta^2}{1-\delta^2(T-h+1)} \left(\sum_{t=h}^T \delta^{t-h} E_t \right)' E[S_E]^{-1} \left(\sum_{t=h}^T \delta^{t-h} E_t \right)$, y,

$$\begin{aligned} L_{T,k,h}(m1_k) &= \frac{E \left[\left(\sum_{t=h}^T \delta^{t-h} E_t \right)' E[S_E]^{-1} \left(\sum_{t=h}^T \delta^{t-h} E_t \right) \right]}{E \left[\left(\sum_{t=h}^T \delta^{t-h} E_t \right)' \left(\sum_{t=h}^T \delta^{t-h} E_t \right) \right]} = \\ &= 1 - \frac{m^2 \frac{\beta_T}{T}}{1 + m^2 \frac{\beta_T}{T} k} \frac{E \left[\left(\sum_{t=h}^T \delta^{t-h} E_t \right)' 1_k 1_k' \left(\sum_{t=h}^T \delta^{t-h} E_t \right) \right]}{E \left[\left(\sum_{t=h}^T \delta^{t-h} E_t \right)' \left(\sum_{t=h}^T \delta^{t-h} E_t \right) \right]} = \\ &= 1 - \frac{m^2 \frac{\beta_T}{T}}{\left(1 + \frac{(1+\delta)(1-\delta^{(T-h+1)})}{(1-\delta)(1+\delta^{(T-h+1)})} m^2 \right)} \frac{\left(1 + \frac{(1+\delta)(1-\delta^{(T-h+1)})}{(1-\delta)(1+\delta^{(T-h+1)})} m^2 k \right)}{(1 + m^2 \frac{\beta_T}{T} k)}. \end{aligned}$$

■

Demostración del Lema 3.3. El coeficiente de curtosis de y_t se puede escribir como:

$$\gamma_y(v) = E \left[\frac{1}{T} \sum_{t=1}^T \left(y_t - \frac{1}{T} \sum_{l=1}^T y_l \right)^4 \right] = \frac{1}{T} \sum_{t=1}^T E \left[\left(y_t - \frac{1}{T} \sum_{l=1}^T y_l \right)^4 \right].$$

Como $y_t = x_t + r_t^{(h,T)}$ y teniendo en cuenta que x_t y $r_t^{(h,T)}$ son independientes y $E[x_t] = E[x_t^3] = 0$, se obtiene:

$$E \left[\left(y_t - \frac{1}{T} \sum_{l=1}^T y_l \right)^4 \right] = E[(x_t + r_t - \bar{r})^4] = E[x_t^4] + 6E[x_t^2](r_t - \bar{r})^2 + (r_t - \bar{r})^4,$$

y,

$$\begin{aligned} \gamma_y(v) &= \frac{1}{T} \sum_{t=1}^T (E[x_t^4] + 6E[x_t^2](r_t - \bar{r})^2 + (r_t - \bar{r})^4) = \\ &= \frac{1}{T} \sum_{t=1}^T E[x_t^4] + \frac{6}{T} \sum_{t=1}^T E[x_t^2](r_t - \bar{r})^2 + \frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^4. \end{aligned}$$

Por último, como $E[x_t^2] = v'v$, $E[x_t^4] = 3E[x_t^2]^2 = 3(v'v)^2$, $\frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^2 = v'\Sigma_R v$ y $v'v = v'\Sigma_Y v - v'\Sigma_R v$, se obtiene,

$$\gamma_y(v) = 3(v'\Sigma_Y v)^2 - 3(v'\Sigma_R v)^2 + \omega_r(v).$$

■

Demostración del Lema 3.4. El Lagrangiano correspondiente para obtener los puntos extremos de $\gamma_y(v)$ está dado por:

$$\mathcal{L}(v) = 3 - 3(v' \Sigma_R v)^2 + \omega_r(v) - \lambda(v' \Sigma_Y v - 1),$$

y su gradiente por:

$$\nabla \mathcal{L}(v) = -12(v' \Sigma_R v) \Sigma_R v + \left(\frac{4}{T} \sum_{t=1}^T (r_t - \bar{r})^2 B_t \right) v - 2\lambda \Sigma_Y v = 0.$$

Multiplicando por v' en $\nabla \mathcal{L}(v)$ y teniendo en cuenta que $v' \Sigma_Y v = 1$, se obtiene el valor $\lambda = -6(v' \Sigma_R v)^2 + 2\omega_r(v)$. Como $\Sigma_R = \frac{1}{T} \sum_{t=1}^T B_t$, entonces,

$$-12(v' \Sigma_R v) \Sigma_R v + 4 \left(\frac{1}{T} \sum_{t=1}^T (v' B_t v) B_t \right) v = \left(-12(v' \Sigma_R v)^2 + \frac{4}{T} \sum_{t=1}^T (v' B_t v)^2 \right) (I + \Sigma_R) v.$$

Por lo tanto,

$$\begin{aligned} -3(v' \Sigma_R v) \Sigma_R v + 3(v' \Sigma_R v)^2 \Sigma_R v + \left(\frac{1}{T} \sum_{t=1}^T (v' B_t v) B_t \right) v - \frac{1}{T} \sum_{t=1}^T (v' B_t v)^2 \Sigma_R v = \\ = -3(v' \Sigma_R v)^2 v + \frac{1}{T} \sum_{t=1}^T (v' B_t v)^2 v, \end{aligned}$$

y, finalmente:

$$\sum_{t=1}^T \left[(v' B_t v) - 3(v' \Sigma_R v) - \frac{\mu(v)}{T} \right] B_t v = T(v' \Sigma_R v)^2 (\gamma_r(v) - 3) v.$$

Las direcciones extremas de $\mathcal{L}(v)$ bajo la restricción $v' \Sigma_Y v = 1$, son los autovectores de la matriz:

$$\left[\sum_{t=1}^T \beta_t(v) B_t \right] v = \mu(v) v,$$

donde $\beta_t(v) = \left[(v' B_t v) - 3(v' \Sigma_R v) - \frac{\mu(v)}{T} \right]$ y $\mu(v) = T(v' \Sigma_R v)^2 (\gamma_r(v) - 3)$. De (3.9), se obtiene que:

$$\gamma_y(v) = 3 - \sigma_r^4(3 - \gamma_r(v)) = 3 + \frac{\mu(v)}{T}.$$

En consecuencia, el máximo y el mínimo de $\gamma_y(v)$ viene dado cuando $\mu(v)$ es lo mayor y lo menor posible, respectivamente. Por lo tanto, el máximo y el mínimo del coeficiente de curtosis coinciden con el máximo y el mínimo de los autovalores de la matriz $\bar{r} = \frac{1}{T} \sum_{t=1}^T \beta_t(v) B_t$.

■

Demostración del Teorema 3.5. En todos los puntos de la demostración, se utilizarán las siguientes igualdades,

$$v' \Sigma_R v = \frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^2 \quad v' B_t v = (r_t - \bar{r})^2 \quad (v' \Sigma_R v)^2 \gamma_r(v) = \frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^4,$$

donde $\bar{r} = \frac{1}{T} \sum_{t=1}^T r_t$.

1. En el caso MAO, $r_h = v'w$, $r_t = 0$, $\forall t \neq h$ y $\bar{r} = \frac{1}{T} r_h$. Primero, $T(v' \Sigma_R v)^2 \gamma_r(v) = c_1 r_h^4$ y $v' \Sigma_R v = c_2 r_h^2$, donde,

$$c_1 = \left(1 - \frac{1}{T}\right) \left[\left(1 - \frac{1}{T}\right)^3 + \frac{1}{T^3} \right] \quad c_2 = \frac{1}{T} \left(1 - \frac{1}{T}\right),$$

y en consecuencia, los autovalores están dados por $\mu(v) = c_0 r_h^4$, donde,

$$c_0 = c_1 - 3T c_2^2 = \left(1 - \frac{1}{T}\right) \left[1 - \frac{6}{T} \left(1 - \frac{1}{T}\right) \right].$$

Por otro lado, después de algunos cálculos se puede demostrar que,

$$\left[\sum_{t=1}^T \beta_t(v) B_t \right] v = [m_1 r_h^3 + m_2 r_h^5] R_h,$$

donde,

$$m_1 = \left(1 - \frac{1}{T}\right) \left[\frac{1}{T^3} + \left(1 - \frac{1}{T}\right)^3 - 3c_2 \right] \quad m_2 = -c_0 \frac{1}{T} \left(1 - \frac{1}{T}\right).$$

Como $R_h = w$,

$$v = \frac{m_1 r_h^3 + m_2 r_h^5}{c_0 r_h^4} w.$$

y los otros autovectores son ortogonales a w . Como los autovalores están dados por $c_0 r_h^4$ y $c_0 > 0$ para $T > 5$, se obtiene que el máximo del coeficiente de curtosis está dado en la dirección de w , mientras que el mínimo se obtiene en las direcciones ortogonales a w .

2. En el caso de un MTC, $r_t = 0$ si $t < h$, $r_t = \delta^{t-h} r_h$ para $t \geq h$ y $\bar{r} = m r_h$, donde $m = (1 - \delta^{T-h+1}) / (T(1 - \delta))$. Primero, $T(v' \Sigma_R v)^2 \gamma_r(v) = c_1 r_h^4$ y $v' \Sigma_R v = c_2 r_h^2$, donde,

$$c_1 = (h-1)m^4 + \sum_{t=h}^T (\delta^{t-h} - m)^4 \quad c_2 = \frac{1}{T} \left[(h-1)m^2 + \sum_{t=h}^T (\delta^{t-h} - m)^2 \right],$$

y consecuentemente, los autovalores están dados por $\mu(v) = c_0 r_h^4$, donde $c_0 = c_1 - 3Tc_2^2$. Por otro lado, se puede demostrar que,

$$\left[\sum_{t=1}^T \beta_t(v) B_t \right] v = [m_1 r_h^3 + m_2 r_h^5] R_h,$$

donde,

$$m_1 = (h-1)(m^4 - 3c_2 m^2) + \sum_{t=h}^T [(\delta^{t-h} - m)^4 - 3c_2 (\delta^{t-h} - m)^2]$$

$$m_2 = -\frac{c_0}{T} \left[(h-1)m^2 + \sum_{t=h}^T (\delta^{t-h} - m)^2 \right],$$

y un autovector es proporcional a w y los otros son ortogonales a él. Como los autovalores están dados por $c_0 r_h^4$, el coeficiente de curtosis de y_t es maximizado o minimizado cuando v es proporcional a w dependiendo del signo de c_0 , que en general depende de los valores de T , h y δ .

3. En el caso MLS, $r_t = 0$ si $t < h$, $r_t = r_h$ para $t \geq h$ y $\bar{r} = \frac{T-h+1}{T} r_h$. Primero, $T(v' \Sigma_R v)^2 \gamma_r(v) = c_1 r_h^4$ y $v' \Sigma_R v = c_2 r_h^2$, donde,

$$c_1 = \frac{(h-1)(T-h+1)}{T^4} [(T-h+1)^3 + (h-1)^3] \quad c_2 = \frac{(h-1)(T-h+1)}{T^2},$$

y los autovalores están dados por $\mu(v) = c_0 r_h^4$, donde,

$$c_0 = c_1 - 3Tc_2^2 = \frac{(h-1)(T-h+1)}{T^3} [T^2 - 6T(h-1) + 6(h-1)^2].$$

Por otro lado, se puede demostrar que,

$$\left[\sum_{t=1}^T \beta_t(v) B_t \right] v = [m_1 r_h^3 + m_2 r_h^5] R_h,$$

donde,

$$m_1 = \frac{(h-1)(T-h+1)}{T^4} [(T-h+1)^3 + (h-1)^3 - 3c_2]$$

$$m_2 = -c_0 \frac{(h-1)(T-h+1)}{T^2},$$

lo que demuestra que un autovector es proporcional a w y los otros son ortogonales a él. Los autovalores están dados por $c_0 r_h^4$ y es fácil ver que,

$$c_0 < 0 \quad \Longleftrightarrow \quad h \in \left(1 + \frac{1}{2} \left(1 - \frac{1}{\sqrt{3}}\right) T, 1 + \frac{1}{2} \left(1 + \frac{1}{\sqrt{3}}\right) T\right)$$

$$c_0 > 0 \quad \Longleftrightarrow \quad h \notin \left(1 + \frac{1}{2} \left(1 - \frac{1}{\sqrt{3}}\right) T, 1 + \frac{1}{2} \left(1 + \frac{1}{\sqrt{3}}\right) T\right).$$

Por lo tanto, el máximo del coeficiente de curtosis se obtiene en la dirección de w si $c_0 > 0$ y el mínimo en la dirección de w si $c_0 < 0$.

■

Demostración del Corolario 3.6. La demostración es inmediata del Teorema 3.5, porque la relación $A_t = E_t + wI_t^{(h)}$ coincide con el caso MAO en una serie de ruido blanco. ■

Capítulo 4

Cambios de varianza en series temporales multivariantes

4.1. Introducción

El problema de detectar un cambio repentino en la varianza marginal de una serie temporal univariante ha sido ampliamente estudiado como se ha comentado en el capítulo 2. Sin embargo, el caso de secuencias multivariantes no ha sido considerado hasta ahora. En este capítulo se estudia la detección de cambios repentinos en la varianza y la estructura de correlación de los componentes de un modelo vector autorregresivo y de media móvil (VARMA). Se introducen y comparan dos puntos de vista. El primero utiliza la razón de verosimilitudes, y puede verse como una generalización del procedimiento univariante propuesto por Tsay (1988). El segundo es un procedimiento cusum, y puede ser visto como una generalización del procedimiento univariante de Inclán y Tiao (1994).

El resto de este capítulo está organizado como sigue. En la sección 4.2, se presenta el modelo para el cambio de varianza y se introducen dos estadísticos para contrastar cambios de varianza cuando los parámetros del modelo VARMA son conocidos. En la sección 4.3, se estudian dos procedimientos para la detección y estimación de estos cambios. En la sección 4.4, se extienden estos procedimientos para permitir cambios en la estructura de correlación. En la sección 4.5 los dos procedimientos se comparan mediante un experimento de Monte Carlo. Por último, en la sección 4.6, los procedimientos se ilustran por medio de dos ejemplos de datos reales. Se concluye que el procedimiento basado en el estadístico cusum tiene mejor comportamiento global que el basado en el contraste de la razón de verosimilitudes.

4.2. Contrastes para cambios de varianza en series multivariantes

Sea $X_t = (X_{1t}, \dots, X_{kt})'$, $t = 1, \dots, T$ un vector de series temporales k -dimensional que sigue un modelo vector ARMA como ha sido definido en (2.7). El modelo de cambio de varianza propuesto por Tsay (1988) se generaliza suponiendo que en lugar de observar X_t se observa la serie temporal $Y_t = (Y_{1t}, \dots, Y_{kt})'$, definida como sigue. Sean $S_t^{(h)}$ una función escalón y W una matriz de constantes diagonal de tamaño $k \times k$ que denota el impacto de un cambio de varianza. Entonces, se supone que las innovaciones que afectan a la serie, E_t , tienen un cambio en la varianza de los componentes en $t = h$, dado por,

$$E_t = A_t + WS_t^{(h)}A_t, \quad (4.1)$$

y, por lo tanto, el vector de series observado, Y_t , se puede escribir como,

$$\Phi(B)Y_t = C + \Theta(B)(A_t + WS_t^{(h)}A_t),$$

y, utilizando (2.7), la relación entre la serie observada, Y_t , y el vector de series ARMA no observado, X_t , está dada por,

$$Y_t = X_t + \Psi(B)WS_t^{(h)}A_t. \quad (4.2)$$

La varianza de E_t cambia de Σ a $\Omega = (I + W)\Sigma(I + W)$ en el tiempo $t = h$. Sin pérdida de generalidad se supone que $(I + W)$ es una matriz definida positiva, por lo que la matriz W está bien identificada. Para ello, las descomposiciones espectrales de las matrices Σ y Ω están dadas por $\Sigma = D_\Sigma R_\Sigma D_\Sigma$ y $\Omega = D_\Omega R_\Omega D_\Omega$ respectivamente, donde R_Σ y R_Ω son las matrices de correlación de Σ y Ω que se suponen iguales, y D_Σ y D_Ω son matrices diagonales cuyos elementos son las desviaciones típicas de cada componente. Entonces, tomando,

$$W = D_\Omega D_\Sigma^{-1} - I, \quad (4.3)$$

se obtiene que $\Omega = (I + W)\Sigma(I + W)$, y la matriz W es única. Se hace notar que el cambio de varianza puede afectar a una o varias de las componentes, de tal manera que los elementos diferentes de 0 de W indican los componentes con cambios.

Para contrastar la significatividad de un cambio de varianza en $t = h$, se supone que los parámetros del modelo ARMA son conocidos y con ellos se pueden obtener los residuos:

$$E_t = Y_t - \sum_{i=1}^p \Phi_i Y_{t-i} - C + \sum_{j=1}^q \Theta_j E_{t-j}. \quad (4.4)$$

Se busca contrastar la hipótesis de que los residuos son homocedásticos e independientes, contra la hipótesis alternativa de que son heterocedásticos. Esto es, se considera la hipótesis nula $H_0 : W = 0$ contra la hipótesis alternativa $H_1 : W \neq 0$. El método más usual para

contrastar la homogeneidad de dos matrices de covarianzas de dos poblaciones Gaussianas es el contraste de la razón de verosimilitud (LRT), que es asintóticamente el test de mayor potencia. Se definen los tres valores $s(i) = \sum_{t=1}^T (E_{it}^2)/T$, $s_1(i) = \sum_{t=1}^{h-1} (E_{it}^2)/(h-1)$ y $s_2(i) = \sum_{t=h}^T (E_{it}^2)/(T-h+1)$. El estadístico de razón de verosimilitudes para un cambio de varianza en el punto $t = h$ está dado por,

$$LR_h = \log \frac{[s(1) \cdots s(k)]^T}{[s_1(1) \cdots s_1(k)]^{h-1} [s_2(1) \cdots s_2(k)]^{T-h+1}}, \quad (4.5)$$

que, bajo la hipótesis nula de que no existe cambio de varianza y de que el modelo es conocido, tiene una distribución asintótica chi cuadrado con k grados de libertad.

Se puede construir un estadístico alternativo como sigue. Bajo la hipótesis nula de homocedasticidad, la matriz de covarianzas de E_t se puede escribir como $\Sigma = D_\Sigma R_\Sigma D_\Sigma$. Se define $B_t = D_\Sigma^{-1} E_t$ con $Cov(B_t) = R_\Sigma$. Los componentes principales de la serie B_t están dados por $C_t = U_\Sigma B_t$, donde U_Σ es la matriz cuyas columnas son los autovectores de la matriz R_Σ , y $Cov(C_t) = \Lambda$, que es una matriz diagonal. Las componentes de C_t son incorreladas con varianzas iguales a los elementos de la diagonal principal de la matriz Λ . Sea $D_m = \sum_{t=1}^m C_t' C_t$ la suma de cuadrados multivariante de la secuencia $\{C_1, \dots, C_m\}$ donde m es cualquier valor $1 \leq m \leq T$. Sea,

$$F_m = \frac{D_m}{D_T} - \frac{m}{T}, \quad m = 1, \dots, T \quad (4.6)$$

donde $F_T = 0$, es la suma de cuadrados acumulada centrada y normalizada de la secuencia C_t . A continuación, se estudia el comportamiento asintótico del estadístico (4.6) bajo la hipótesis de homocedasticidad. Las demostraciones del Lema 4.1 y del Teorema 4.2, enunciados a continuación, aparecen en el apéndice de este capítulo.

Lema 4.1 *Si la matriz de covarianzas de la secuencia $\{E_1, \dots, E_T\}$ en (4.4) es constante, para un valor dado $t = m$,*

$$E[F_m] = o(T^{-1}).$$

Consecuentemente, el valor esperado del estadístico F_m de 4.6 es asintóticamente 0 para todo m . A continuación, se estudia la distribución asintótica del estadístico F_m bajo la hipótesis de que no existe cambio para $t = 1, \dots, T$. Sea M un movimiento Browniano que verifica $E[M_r] = 0$, y $E[M_r M_s] = s$, donde $0 \leq s < r \leq 1$. Sea M^0 un puente Browniano dado por $M_r^0 = M_r - rM_1$, verificando $E[M_r^0] = 0$, $E[M_r^0 M_s^0] = s(1-r)$, $0 \leq s < r \leq 1$, y $M_0^0 = M_1^0 = 0$, con probabilidad 1. La distribución asintótica del estadístico F_m se obtiene en el siguiente Teorema.

Teorema 4.2 *Sea $\{E_1, \dots, E_T\}$ una secuencia de variables aleatorias Gaussianas independientes e igualmente distribuidas de media 0 y matriz de covarianzas común Σ . Sea $B_t = D_\Sigma^{-1} E_t$ con $Cov(B_t) = \Lambda$, donde D_Σ es una matriz diagonal cuyos elementos son las desviaciones típicas de las componentes de E_t y sean C_t los componentes principales de la serie B_t .*

Si $F_m = D_m/D_T - m/T$, donde $D_m = \sum_{t=1}^m C_t' C_t$, entonces, $D_m^* = \sqrt{\frac{T}{2}} \frac{k}{\sqrt{\text{tr}(\Lambda^2)}} F_m \xrightarrow{D} M^0$, donde tr denota traza.

Se ha demostrado que la distribución asintótica del estadístico D_m^* bajo la hipótesis de matriz de covarianzas constante es un puente Browniano. Por ello, se puede utilizar el estadístico D_{h-1}^* para contrastar la presencia de un cambio en la matriz de covarianza en $t = h$, y el valor crítico asintótico de la distribución de un puente Browniano. El estadístico D_m^* depende del valor de $\text{tr}(\Lambda^2)$, que en la práctica es desconocido. Sea $\Lambda(i, i)$ el elemento íésimo de la diagonal de la matriz Λ . Bajo la hipótesis de no cambio, podemos estimar $\Lambda(i, i)$ por medio de $\hat{\Lambda}(i, i) = \sum_{t=1}^T (C_{it}^2)/T$, que es un estimador consistente de $\Lambda(i, i)$. Si $\hat{\Lambda}_1^m(i, i) = \sum_{t=1}^m (C_{it}^2)/m$ y teniendo en cuenta que $\text{tr}(\hat{\Lambda}) = k$, entonces, se define el estadístico G_m , como sigue,

$$G_m = \sqrt{\frac{1}{2T}} \frac{mk}{\sqrt{\text{tr}(\hat{\Lambda}^2)}} \left(\frac{\hat{\Lambda}_1^m(1, 1) + \cdots + \hat{\Lambda}_1^m(k, k)}{k} - 1 \right). \quad (4.7)$$

Bajo la hipótesis de no existencia de cambio en las varianzas, puesto que $\hat{\Lambda}(i, i)$ es un estimador consistente de $\Lambda(i, i)$, los estadísticos D_m^* y G_m tienen la misma distribución asintótica.

El impacto de un cambio de varianza se puede estimar como sigue. Sean $\Omega(i, i)$, $\Sigma(i, i)$ y $W(i, i)$ los elementos íésimos de las diagonales principales de las matrices Ω , Σ and W , respectivamente. Entonces,

$$(1 + W(i, i))^2 = \frac{\Omega(i, i)}{\Sigma(i, i)}, \quad i = 1, \dots, k$$

Se definen las matrices $S_1 = \sum_{t=1}^{h-1} (E_t E_t') / (h-1)$ y $S_2 = \sum_{t=h}^T (E_t E_t') / (T-h+1)$ y como los estimadores máximo verosímiles de Σ y Ω están dados por $\hat{\Sigma} = S_1$ y $\hat{\Omega} = S_2$, se estima $\widehat{W}(i, i)$ por:

$$(1 + \widehat{W}(i, i))^2 = \frac{s_2(i)}{s_1(i)}, \quad i = 1, \dots, k$$

dado que $s_1(i)$ y $s_2(i)$ son los elementos íésimos de las diagonales principales de las matrices S_1 y S_2 , respectivamente. Bajo la hipótesis nula de no existencia de cambio de varianzas, $(1 + \widehat{W}(i, i))^2$ sigue una distribución F con $(T-h, h-2)$ grados de libertad. Por lo tanto, se puede contrastar la hipótesis nula $W(i, i) = 0$ contra la alternativa de ser diferentes por medio de la distribución F . Como $1 + \widehat{W}(i, i)$ debe ser mayor que 1, se pueden obtener estimaciones finales de $\widehat{W}(i, i)$ mediante:

$$\widehat{W}(i, i) = \sqrt{\frac{s_2(i)}{s_1(i)}} - 1. \quad i = 1, \dots, k \quad (4.8)$$

Un intervalo de confianza para $W(i, i)$ para un nivel de significatividad α está dado por:

$$1 - \alpha = P \left(\frac{1 + \widehat{W}(i, i)}{\sqrt{F_{(T-h, h-2)}^{1-\alpha/2}}} - 1 \leq W(i, i) \leq \frac{1 + \widehat{W}(i, i)}{\sqrt{F_{(T-h, h-2)}^{\alpha/2}}} - 1 \right)$$

donde $F_{(T-h, h-2)}^{\alpha/2}$ y $F_{(T-h, h-2)}^{1-\alpha/2}$ son los valores críticos de la distribución F con $(T - h, h - 2)$ grados de libertad para los niveles de significatividad $\alpha/2$ y $1 - \alpha/2$, respectivamente.

4.3. Procedimientos para la detección de cambios de varianza

Una serie puede estar afectada por varios cambios de varianza. En este caso, se observa una serie temporal Y_t , definida como sigue:

$$Y_t = X_t + \Psi(B)(I + W_r S_t^{(h_r)}) \cdots (I + W_1 S_t^{(h_1)}) A_t,$$

donde $\{h_1, \dots, h_r\}$ son los instantes de los r puntos de cambio y W_1, \dots, W_r son matrices diagonales de tamaño $k \times k$ que denotan el impacto de los r cambios. Suponiendo que los parámetros son conocidos, la serie filtrada de residuos está dada por:

$$E_t = (I + W_r S_t^{(h_r)}) \cdots (I + W_1 S_t^{(h_1)}) A_t,$$

y la matriz de covarianzas residual de E_t cambia de Σ a $(I + W_1) \Sigma (I + W_1)$ en $t = h_1$, a $(I + W_2)(I + W_1) \Sigma (I + W_1)(I + W_2)$ en $t = h_2, \dots$

En la práctica, los parámetros de un modelo VARMA, el número, la localización y los tamaños de los cambios de varianza son desconocidos. Sean $\widehat{L}R_t$ y \widehat{G}_t los estadísticos (4.5) y (4.7) respectivamente calculados usando los residuos estimados que son obtenidos mediante (4.4). Se definen los máximos de estos estadísticos en la muestra como,

$$\Lambda_{\max}(h_{\max}^{LR}) = \max \left\{ \left| \widehat{L}R_t \right|, 1 \leq t \leq T \right\}, \quad \Gamma_{\max}(h_{\max}^G) = \max \left\{ \left| \widehat{G}_t \right|, 1 \leq t \leq T \right\}, \quad (4.9)$$

donde h_{\max}^{LR} y $h_{\max}^G + 1$ son las estimaciones de los tiempos de cambio utilizando el estadístico LRT o el estadístico cusum, respectivamente. La distribución de Λ_{\max} en (4.9) es intratable y valores críticos deben ser obtenidos mediante simulación. La distribución de Γ_{\max} en (4.9) es asintóticamente la distribución de $\sup \{|M_r^0| : 0 \leq r \leq 1\}$ que está dada por (ver, Billingsley, pg. 85, 1968),

$$P \left\{ \sup |M_r^0| \leq a : 0 \leq r \leq 1 \right\} = 1 + 2 \sum_{i=1}^{\infty} (-1)^i \exp(-2i^2 a^2),$$

de donde pueden ser obtenidos valores críticos. A continuación se proponen dos algoritmos iterativos para detectar varios cambios y estimar sus impactos basados en los estadísticos Λ_{\max} and Γ_{\max} .

Para motivar los procedimientos propuestos supongamos una serie bivalente de 500 observaciones de un modelo vectorial autorregresivo de primer orden. Se consideran las tres situaciones ilustradas en la Figura 4.1. Las tres columnas en esta matriz de gráficos representan tres procesos generadores diferentes. La primera columna corresponde al caso en que no hay cambios de varianza. La segunda columna corresponde al caso de un cambio en la matriz de covarianzas en $t = 250$, donde la matriz de covarianzas de las innovaciones cambia de I a la matriz $3 \times I$. La tercera columna corresponde al caso de dos cambios en $t = 166$, donde la matriz de covarianzas de las innovaciones cambia de I a $3 \times I$, y en $t = 333$, donde la matriz de covarianzas vuelve a ser I . Las filas representan las series de las dos componentes de la serie temporal bivalente y la de los dos estadísticos introducidos en la sección previa. La primera (segunda) fila en la Figura 4.1 muestra la primera (segunda) componente de esta serie bivalente. La tercera y cuarta filas muestran las series de valores del estadístico de razón de verosimilitudes (4.5) y del estadístico cusum (4.7) respectivamente. Estos estadísticos se han calculado con la serie bivalente presentada en las dos primeras filas de la misma columna. El caso de varianza constante corresponde a la primera columna de la Figura 4.1. En este caso los dos estadísticos dibujados en la tercera y cuarta filas están dentro de las dos líneas rectas, calculadas como será explicado más adelante, que representan el intervalo de confianza al nivel del 95 % de las distribuciones de Λ_{\max} y Γ_{\max} . La segunda columna corresponde al caso de un cambio en $t = 250$. El máximo de ambos estadísticos en valor absoluto está alrededor de $t = 250$, y es mayor que el valor crítico correspondiente, por lo que se rechaza la hipótesis de varianza constante. La tercera columna corresponde a dos cambios de nivel y existen dos extremos significativos alrededor de los tiempos de los cambios $t = 166$ y $t = 333$.

4.3.1. Procedimiento LRT

Especificar un modelo vectorial ARMA para la serie observada Y_t y obtener las estimaciones máximo verosímiles de los parámetros del modelo así como la serie de residuos.

1. Calcular los estadísticos LR_h , $h = H + 1, \dots, T - H$, para un valor dado de un entero H , utilizando los residuos obtenidos en el paso 1. El número $H = k + m$ es un entero positivo que denota el mínimo número de residuos necesarios para estimar la matriz de covarianzas. El valor m puede ser fijado por el usuario, y en los ejemplos y experimentos de Monte Carlo se ha tomado $m = 10$. Con ellos, se obtiene el estadístico $\Lambda_{\max}(h_{\max}^{LR})$ en (4.9).
2. Comparar $\Lambda_{\max}(h_{\max}^{LR})$ con un valor crítico específico D para un nivel crítico dado. Si $\Lambda_{\max}(h_{\max}^{LR}) < D$, se concluye que no existe un cambio de varianzas y el procedimiento

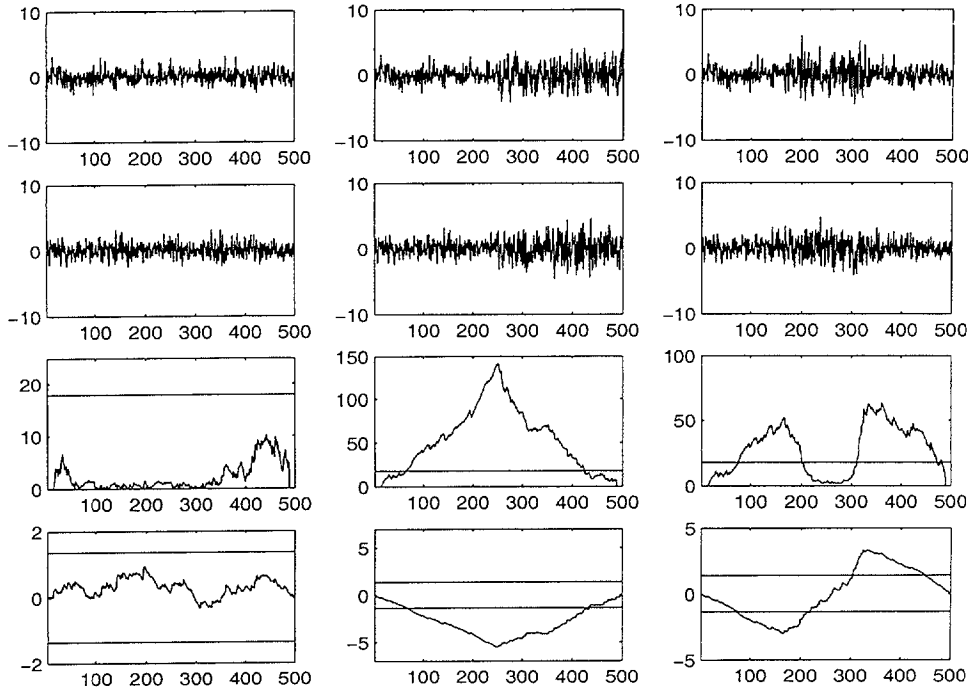


Figura 4.1: Serie bivalente y estadísticos para detección de cambio de varianza.

termina. Si $\Lambda_{\max}(h_{\max}^{LR}) \geq D$, se asume que se detecta un cambio de varianzas en tiempo $t = h_{\max}^{LR}$.

3. Se estima la matriz \widehat{W} por (4.8) y se obtiene una serie de residuos modificada como sigue:

$$E_t^* = \begin{cases} \widehat{E}_t & t < h_{\max}^{LR} \\ (I + \widehat{W})^{-1} \widehat{E}_t & t \geq h_{\max}^{LR}, \end{cases}$$

y, con esta serie de residuos, se define una serie corregida por,

$$Y_t^* = \begin{cases} Y_t & t < h_{\max}^{LR} \\ \widehat{C} + \widehat{\Phi}_1 Y_{t-1}^* + \dots + \widehat{\Phi}_p Y_{t-p}^* + E_t^* - \widehat{\Theta}_1 E_{t-1}^* - \dots - \widehat{\Theta}_q E_{t-q}^* & t \geq h_{\max}^{LR}, \end{cases}$$

donde los polinomios $\widehat{\Phi}(B)$ y $\widehat{\Theta}(B)$ son los estimadores máximo verosímiles de los parámetros. Entonces, volver al Paso 1 considerando Y_t^* como el proceso observado.

4. Cuando no se detecten más cambios de varianza, los parámetros de la serie y todos los cambios de varianza detectados en los pasos previos se estiman conjuntamente, mediante el modelo,

$$\Phi(B)Y_t = C + \Theta(B)(I + W_r S_t^{(h_r)}) \dots (I + W_1 S_t^{(h_1)}) A_t. \quad (4.10)$$

Esta estimación conjunta se lleva a cabo en dos pasos. Primero, se estiman los parámetros suponiendo que no existen cambios de varianza y segundo con los residuos se estiman las matrices W_i . Después se corrige la serie, y se repiten estos dos pasos hasta convergencia.

4.3.2. Procedimiento cusum

El siguiente procedimiento es una generalización del propuesto por Inclán y Tiao (1994). El algoritmo está basado en sucesivas divisiones de la serie en dos partes cuando se detecta un cambio y procede como sigue:

1. Especificar un modelo ARMA vectorial para la serie observada Y_t , suponiendo que no hay cambios de varianza, y obtener las estimaciones máximo verosímiles así como la serie de residuos. A continuación, obtener los componentes principales de la serie de residuos, C_t , como en la sección 4.2. Se define $t_1 = 1$.
2. Obtener $\Gamma_{\max}(h_{\max}^G)$ para C_t en (4.9) para $t = 1, \dots, T$. Si $\Gamma_{\max}(h_{\max}^G) > D$, donde D es el valor crítico asintótico para un nivel crítico dado, ir al Paso 3. Si $\Gamma_{\max}(h_{\max}^G) < D$, se asume que no existe un cambio de varianza en la serie y el procedimiento finaliza.
3. El Paso 3 tiene tres subpasos:
 - a) Obtener $\Gamma_{\max}(h_{\max}^G)$ para $t = 1, \dots, t_2$, donde $t_2 = h_{\max}^G$. Si $\Gamma_{\max}(h_{\max}^G) > D$, redefinir $t_2 = h_{\max}^G$ y repetir el Paso 3(a) hasta que $\Gamma_{\max}(h_{\max}^G) < D$. Cuando esto ocurra, definir $h_{first} = t_2$ donde t_2 es el último valor tal que $\Gamma_{\max}(h_{\max}^G) > D$.
 - b) Repetir una búsqueda similar en el intervalo $t_2 \leq t \leq T$, donde t_2 es el punto h_{\max}^G obtenido en el Paso 2. Para ello, definir $t_1 = h_{\max}^G + 1$, donde $h_{\max}^G = \arg \max \{G_t : t = t_1, \dots, T\}$ y repetir hasta que $\Gamma_{\max}(h_{\max}^G) < D$. Definir $h_{last} = t_1 - 1$, donde t_1 es el último valor tal que $\Gamma_{\max}(h_{\max}^G) > D$.
 - c) Si $|h_{last} - h_{first}| < H$, existe un único punto de cambio y el algoritmo finaliza aquí. Si no, mantener ambos valores como posibles puntos de cambios y repetir los pasos 2 y 3 para $t_1 = h_{first}$ y $T = h_{last}$, hasta que no se detecten más posibles puntos de cambio. Entonces, ir al paso 4.
4. Definir un vector $\ell = (\ell_1, \dots, \ell_s)$ donde $\ell_1 = 1$, $\ell_s = T$ y $\ell_2, \dots, \ell_{s-1}$ son los puntos detectados en los pasos 2 y 3 en orden creciente. Obtener el estadístico G_t en cada uno de los intervalos (ℓ_i, ℓ_{i+2}) y comprobar si los valores máximos son todavía significativos. Si alguno no lo es, se elimina el correspondiente punto. Repetir el paso 4 hasta que el número de posibles cambios no varíe, y los puntos encontrados en las iteraciones anteriores no difiera de la última. El vector $(\ell_2 + 1, \dots, \ell_{s-1} + 1)$ contiene los puntos de cambios de varianza.

5. Por último, se estiman los parámetros de la serie y los cambios de varianza detectados en los pasos anteriores conjuntamente mediante (4.10).

Observemos que: (1) los valores críticos en el algoritmo LRT deben ser obtenidos mediante simulación, como se verá en la siguiente sección, mientras que los valores críticos utilizados en el procedimiento cusum son los valores críticos asintóticos del máximo del valor absoluto de un puente Browniano. (2) En ambos algoritmos se requiere una distancia mínima entre cambios de varianza mayor que H , de tal manera que la matriz de covarianzas pueda ser estimada. Si se encuentran varios cambios en un intervalo menor que H , estos cambios serán considerados como atípicos y estimados mediante el procedimiento descrito en el capítulo 3. (3) El último paso en el procedimiento LRT es necesario para eliminar sesgos en el tamaño de los cambios de varianza estimados. Como en el paso 4, el tamaño del cambio de varianza se estima tras detectarlo, si existen dos o más cambios, el impacto del primer cambio detectado se ha estimado sin tener en cuenta el resto. Entonces, es necesaria una estimación conjunta con todos los cambios detectados.

4.4. Cambios en las varianzas y en las correlaciones

A continuación, se supone que en lugar de X_t se observa una serie temporal $Y_t = (Y_{1t}, \dots, Y_{kt})'$, definida como sigue. Sea W una matriz constante triangular inferior de tamaño $k \times k$. Se supone que las innovaciones que afectan a la serie, E_t , tienen un cambio en la matriz de covarianzas de las componente en el mismo punto $t = h$ dado por (4.1) donde W es triangular inferior y la serie observada Y_t se puede escribir como en (4.2). La matriz de covarianzas de E_t cambia en el punto $t = h$ de Σ a $\Omega = (I + W)\Sigma(I + W)'$. Sin pérdida de generalidad, se supone que $(I + W)$ es una matriz definida positiva, para que la matriz W esté bien identificada. Para ello, sean $\Sigma = L_\Sigma L_\Sigma'$ y $\Omega = L_\Omega L_\Omega'$ las descomposiciones de Cholesky de Σ y Ω , respectivamente. Entonces, tomando,

$$W = L_\Omega L_\Sigma^{-1} - I, \quad (4.11)$$

con W triangular inferior, se obtiene $\Omega = (I + W)\Sigma(I + W)'$, y puesto que la descomposición de Cholesky es única, la matriz W es también única.

Como en el caso anterior, para contrastar la significatividad de un cambio en $t = h$, se supone que los parámetros del modelo vector ARIMA son conocidos y mediante ellos se pueden obtener los residuos como en (4.4). Se considera la hipótesis nula $H_0 : W = 0$ contra la hipótesis alternativa $H_1 : W \neq 0$ y se define la matriz $S = \sum_{t=1}^T (E_t E_t') / T$. El estadístico de la razón de verosimilitudes de los residuos en (4.4) para un cambio en la matriz de covarianzas en el punto $t = h$ está dado por,

$$LR_h = \log \frac{|S|^T}{|S_1|^{h-1} |S_2|^{T-h+1}} \quad (4.12)$$

y bajo la hipótesis nula de que no existe cambio y suponiendo el modelo es conocido, el estadístico LR_h tiene una distribución asintótica χ^2 con $\frac{1}{2}k(k+1)$ grados de libertad.

Un estadístico cusum alternativo se obtiene como sigue. Sea $D_m = \sum_{t=1}^m E_t' \Sigma^{-1} E_t$ la suma de cuadrados acumulada multivariante de $\{E_1, \dots, E_m\}$ donde m es cualquier valor dado $1 \leq m \leq T$. Sea,

$$F_m = \frac{D_m}{D_T} - \frac{m}{T}, \quad m = 1, \dots, T \quad (4.13)$$

donde $F_T = 0$, es la suma de cuadrados acumulada de la secuencia E_t centrada y normalizada. La distribución asintótica del estadístico (4.13) bajo la hipótesis de homocedasticidad se puede obtener con los siguientes Lema y Teorema, cuyas demostraciones son similares al caso de cambio de varianza, por lo que no se presentan.

Lema 4.3 Si la matriz de covarianzas de la secuencia $\{E_1, \dots, E_T\}$ en (4.4) es constante, para un valor dado $t = m$,

$$E[F_m] = o(T^{-1}).$$

La distribución asintótica del estadístico F_m se obtiene en el siguiente Teorema.

Teorema 4.4 Sea $\{E_1, \dots, E_T\}$ una secuencia de variables aleatorias Gaussianas independientes e idénticamente distribuidas con media 0 y matriz de covarianzas común Σ . Si $F_m = D_m/D_T - m/T$, donde $D_m = \sum_{t=1}^m E_t' \Sigma^{-1} E_t$, entonces, el estadístico $F_m^* = \sqrt{Tk/2} F_m \xrightarrow{D} M^0$.

La distribución asintótica del estadístico $F_m^* = (Tk/2)^{\frac{1}{2}} F_m$ bajo la hipótesis de que no existe cambio de covarianzas es un puente Browniano: De este modo, el estadístico F_{h-1}^* puede ser utilizado para contrastar la presencia de un cambio en $t = h$, mediante el valor crítico asintótico de esta distribución. El estadístico F_m^* depende de la matriz de covarianzas Σ , que en la práctica puede ser estimada consistentemente por la matriz $S = \sum_{t=1}^T (E_t E_t') / T$. Se define el estadístico G_m como,

$$G_m = \sqrt{\frac{k}{2T}} m \left(\frac{\text{tr}(S^{-1} S_1^m)}{k} - 1 \right), \quad (4.14)$$

donde $S_1^m = \sum_{t=1}^m (E_t E_t') / m$. Bajo la hipótesis de no existencia de cambio, los estadísticos F_m^* y G_m tienen la misma distribución asintótica.

El impacto de la matriz de covarianzas en (4.11) se estima con,

$$\widehat{W} = L_{S_2} L_{S_1}^{-1} - I. \quad (4.15)$$

Cuando existen varios cambios, los procedimientos LRT y cusum para cambios de varianza se adaptan al caso de cambios en la matriz de covarianzas, donde la matriz W se estima con (4.15). Los estadísticos máximos en (4.9) se definen de la misma manera para los estadísticos (4.12) y (4.14).

Tabla 4.1: Modelos para el estudio de simulación.

$k = 2$				$k = 3$			
Φ		Σ		Φ		Σ	
0.6	0.2	1	0	0.6	0.2	0	1
0.2	0.4	0	1	0.2	0.4	0	0
				0.6	0.2	0.5	0

4.5. Experimentos de Monte Carlo

En esta sección se lleva a cabo un estudio de simulación dividido en varias partes. En primer lugar, se obtienen valores críticos para el estadístico Λ_{\max} en (4.9) para W diagonal y W triangular inferior para dos modelos vector AR(1) con $k = 2, 3$ respectivamente y que aparecen en la Tabla 4.1, y tamaños muestrales $T = 100, 200, 500$ y 1000 . Para cada modelo y tamaño muestral, se han generado 10000 series y se ha estimado un modelo vector AR(1), obtenido los residuos, \hat{E}_t , y calculado los estadísticos (4.9). La Tabla 4.2 proporciona algunos cuantiles de la distribución de Λ_{\max} para ambos modelos y tamaños muestrales bajo la hipótesis de homocedasticidad de la serie. Se hace notar que los cuantiles dependen de la dimensión de la serie temporal. La distribución asintótica del estadístico Γ_{\max} es conocida, pero también se estudia el comportamiento en muestras finitas de los cuantiles de estos estadísticos que se muestran en la Tabla 4.2. Se puede observar que los cuantiles en muestras finitas son siempre menores que los asintóticos, lo que implica que el uso de los cuantiles asintóticos es una decisión conservadora y por lo tanto el error de tipo I no aumentará. En este caso, los cuantiles muestrales no dependen de la dimensión k .

A continuación, se realiza un estudio de simulación del tamaño y la potencia de los procedimientos para cambios de varianza. Para ello, se consideran los modelos de la Tabla 4.1 para $T = 100, 200$ y 500 . Para el caso de un cambio de varianzas, se consideran tres localizaciones del punto de cambio, $h = [0.25T], [0.50T]$ y $[0.75T]$. Estos cambios transforman la matriz de covarianzas original, $\Sigma = I$, en $\Omega = (I + W)(I + W)$, donde W es una matriz diagonal. Se consideran tres posibles matrices de la forma $W = \alpha I$, donde α toma tres posibles valores: $\alpha = 0$, en el caso de no haber cambio de varianzas, $\alpha = \sqrt{2} - 1$, en la que la matriz de covarianzas es multiplicada por 2, y $\alpha = \sqrt{3} - 1$, en el que la matriz de covarianzas es multiplicada por 3. En cada caso, se generan 5000 series y se aplican los dos procedimientos para cada serie con el valor crítico al 95 % de la Tabla 4.2. Los resultados se muestran en las Tablas 4.3 y 4.4, donde las columnas 4 a 6 y 9 a 11 dan el número de cambios de varianza detectados por los algoritmos y las columnas 7, 8, 12 y 13 muestran la mediana y la desviación absoluta media de las estimaciones del punto de cambio en cada caso. Los casos con $\alpha = 0$ indican el error de tipo I de los procedimientos, que es alrededor del 5 % en todos los tamaños muestrales considerados. De estas dos tablas se concluye que

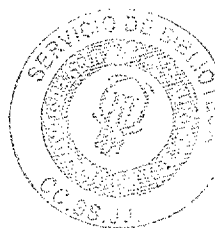


Tabla 4.2: Cuantiles empíricos para los estadísticos Λ_{\max} y Γ_{\max} basados en 10000 series.

W diagonal										
k=2	Probabilidad-LRT					Probabilidad-Cusum				
	50 %	90 %	95 %	97.5 %	99 %	50 %	90 %	95 %	97.5 %	99 %
n=100	9.02	13.94	16.08	17.96	20.02	0.73	1.12	1.27	1.41	1.55
n=200	9.74	15.10	16.76	18.80	21.27	0.76	1.13	1.28	1.42	1.56
n=500	10.87	15.91	17.82	19.92	21.70	0.80	1.18	1.31	1.46	1.61
n=1000	11.24	16.65	18.55	20.27	22.65	0.81	1.19	1.33	1.47	1.62
n= ∞	-	-	-	-	-	0.82	1.22	1.35	1.48	1.62
k=3	50 %	90 %	95 %	97.5 %	99 %	50 %	90 %	95 %	97.5 %	99 %
n=100	11.97	17.76	19.82	21.55	24.68	0.72	1.08	1.22	1.32	1.45
n=200	13.03	18.89	20.79	22.44	25.04	0.77	1.17	1.31	1.41	1.56
n=500	14.34	20.10	22.16	24.13	26.06	0.80	1.19	1.32	1.43	1.57
n=1000	14.71	20.47	22.40	24.40	26.68	0.81	1.20	1.34	1.45	1.59
n= ∞	-	-	-	-	-	0.82	1.22	1.35	1.48	1.62
W triangular inferior										
k=2	Probabilidad-LRT					Probabilidad-Cusum				
	50 %	90 %	95 %	97.5 %	99 %	50 %	90 %	95 %	97.5 %	99 %
n=100	11.07	16.87	18.93	20.67	23.66	0.75	1.16	1.30	1.39	1.55
n=200	11.73	17.29	19.34	21.57	23.97	0.77	1.17	1.31	1.41	1.57
n=500	12.69	18.36	20.60	22.31	25.07	0.79	1.18	1.32	1.42	1.58
n=1000	13.16	19.01	21.44	23.56	26.52	0.80	1.19	1.33	1.44	1.60
n= ∞	-	-	-	-	-	0.82	1.22	1.35	1.48	1.62
k=3	50 %	90 %	95 %	97.5 %	99 %	50 %	90 %	95 %	97.5 %	99 %
n=100	17.37	24.66	27.30	29.15	32.32	0.75	1.13	1.26	1.38	1.51
n=200	18.24	24.94	27.38	29.18	32.68	0.76	1.14	1.29	1.44	1.54
n=500	19.26	25.88	28.47	30.65	34.78	0.78	1.17	1.33	1.44	1.56
n=1000	19.96	26.86	28.86	30.80	34.95	0.80	1.20	1.34	1.46	1.60
n= ∞	-	-	-	-	-	0.82	1.22	1.35	1.48	1.62

Tabla 4.3: Resultados para el modelo 1 y un cambio de varianza.

Ω	T	h	Procedimiento LRT					Procedimiento Cusum				
			frecuencia			\hat{h}		frecuencia			\hat{h}	
			0	1	≥ 2	Med.	Mad	0	1	≥ 2	Med.	Mad
I	100	—	95.8	4.2	0	—	—	96.6	3.4	0	—	—
I	200	—	95.4	4.4	0.2	—	—	95.6	4.2	0.2	—	—
I	500	—	95.6	4.2	0.2	—	—	96.2	3.6	0.2	—	—
$2 \times I$	100	25	55.7	43.5	0.8	26	4	52.3	47.1	0.6	34	7
		50	37.7	61.5	0.8	51	5	14.2	84.8	1.0	52	3
		75	49.5	50.3	0.2	75	4	25.3	74.7	0	74	3
$2 \times I$	200	50	22.8	75.0	2.2	51	6	13.4	83.2	3.4	57	7
		100	7.6	90.6	1.8	100	5	1.2	96.2	2.6	101	3
		150	14.8	84.2	1.0	150	4	4.8	93.8	1.4	150	4
$2 \times I$	500	125	0	98.6	1.4	126	4	0.2	95.4	4.4	130	6
		250	0	96.8	3.2	251	3	0	94.0	6.0	252	3
		375	0.2	97.4	2.4	376	4	0	95.2	4.8	375	4
$3 \times I$	100	25	9.6	88.8	1.6	25	2	9.4	88.4	2.2	28	3
		50	2.2	97.0	0.8	50	2	0.2	98.0	1.8	51	1
		75	6.4	92.6	1.0	75	2	1.6	97.0	1.4	75	2
$3 \times I$	200	50	0	97.4	2.6	50	2	0	95.8	4.2	52	2
		100	0	96.8	3.2	100	2	0	96.6	3.4	101	1
		150	0.2	97.4	2.4	150	1	0.2	95.6	4.2	150	1
$3 \times I$	500	125	0	95.4	4.6	125	1	0	92.0	8.0	126	2
		250	0	97.2	2.8	250	1	0.2	91.8	8.0	251	1
		375	0	96.6	3.4	375	2	0	93.4	6.6	375	1

cuando $T = 100$ el procedimiento cusum funciona mejor que el procedimiento LRT. Para $T = 200$ y para un cambio pequeño, $\alpha = 2$, el procedimiento cusum es ligeramente superior al LRT, pero para un cambio grande $\alpha = 3$, el procedimiento LRT es ligeramente más potente. Las estimaciones del momento de cambio, h , son similares para ambos procedimientos. Para $T = 500$ el algoritmo LRT detecta más frecuentemente un cambio que el cusum, que tiende a detectar más de un cambio en algunas series.

Para dos cambios, se consideran los mismos tamaños muestrales y puntos de cambio en $(h_1, h_2) = ([0.33T], [0.66T])$. Cada punto de cambio es asociado con dos matrices, Ω_1 y Ω_2 , que dan las matrices de covarianzas residuales después del momento de cambio. Se consideran seis combinaciones de estas matrices. Para cada caso, se generan 5000 series con los correspondientes cambios. Entonces, se aplican ambos procedimientos con el valor crítico al 95 % de la Tabla 4.2. Los resultados aparecen en las Tablas 4.5 y 4.6. Las columnas 6 a 9 de esas tablas son los números de cambios de varianzas detectado por cada algoritmo, y las

Tabla 4.4: Resultados para el modelo 2 y un cambio de varianza.

Ω	T	h	Procedimiento LRT					Procedimiento Cusum				
			frecuencia			\hat{h}		frecuencia			\hat{h}	
			0	1	≥ 2	Med.	Mad	0	1	≥ 2	Med.	Mad
I	100	—	95.8	4.0	0.2	—	—	96.0	4.0	0	—	—
I	200	—	93.4	6.2	0.4	—	—	95.4	4.4	0.2	—	—
I	500	—	97.0	3.0	0	—	—	95.8	3.8	0.4	—	—
$2 \times I$	100	25	39.7	59.7	0.6	25	3	30.5	68.1	1.4	30	5
		50	23.4	76.0	0.6	50	4	5.4	92.6	2.0	51	2
		75	41.1	58.3	0.6	75	4	14.4	84.8	0.8	75	3
$2 \times I$	200	50	8.4	90.4	1.2	50	4	1.2	96.2	2.6	53	4
		100	1.2	96.6	2.2	100	3	0.2	97.2	2.6	101	2
		150	5.4	92.8	1.8	150	3	1.2	96.8	2.0	150	2
$2 \times I$	500	125	0	97.8	2.2	125	2	0	92.0	8.0	127	4
		250	0	97.8	2.2	250	2	0.2	94.4	5.4	251	2
		375	0	98.2	1.8	375	3	0.2	92.6	7.2	375	2
$3 \times I$	100	25	1.2	97.6	1.2	25	1	0.8	97.8	1.4	26	2
		50	0	98.2	1.8	50	1	0	96.4	3.6	50	1
		75	2.4	96.4	1.2	75	1	0.4	96.8	2.8	75	1
$3 \times I$	200	50	0	98.4	1.6	50	1	0	96.0	4.0	51	1
		100	0	95.6	4.4	100	1	0.2	92.2	7.6	101	1
		150	0	97.6	2.4	150	1	0	97.0	3.0	150	1
$3 \times I$	500	125	0	97.8	2.2	125	1	0.2	93.0	6.8	126	2
		250	0	96.4	3.6	250	1	0	92.8	7.2	251	1
		375	0	97.8	2.2	375	1	0	95.2	4.8	375	1

columnas 10 a 13 muestran la mediana y la desviación absoluta media de las estimaciones de los puntos de cambio. Para dos puntos de cambio, las ventajas del procedimiento cusum sobre el LRT es clara. En primer lugar la frecuencia de detección de dos puntos de cambio es mayor para el procedimiento cusum en casi todos los casos, y segundo, el procedimiento LRT sufre de una sobreestimación del número de cambios en algunas situaciones. En general, excepto cuando $k = 2$, el tamaño muestral es pequeño ($T = 100$) y pequeños cambios ($\Omega_1 = 2 \times I$, $\Omega_2 = I$), la frecuencia de detección es baja: 17.2 % y 34.9 % para $k = 2$ y $k = 3$, respectivamente. En el resto de casos, el procedimiento cusum funciona bastante bien, con varios casos sobre el 90 % de potencia. Como en el caso previo, cuando el tamaño muestral crece, el cambio es mayor y el número de componentes crece, el procedimiento funciona mejor. También parece que la estimación del segundo cambio tiene menor mad, lo que sugiere que el procedimiento detecta más eficientemente el cambio al final de la serie. La mediana de las estimaciones son bastante aproximadas a los puntos de cambio reales excepto con el menor tamaño muestral y el menor de los cambios.

A continuación se estudia el caso de cambios de varianzas y correlaciones, para lo que se realiza un estudio de simulación que estime la potencia de los procedimientos para el caso de un cambio. Para ello, se consideran los mismos modelos y tamaños muestrales para $k = 2$ que en el caso previo. Los cambios se introducen transformando la matriz de covarianzas original, $\Sigma = I$, en $\Omega = (I + W)(I + W)'$, donde W es una matriz triangular inferior. Se consideran dos posibles matrices W , asociadas con dos matrices Ω_1 y Ω_2 que representan la situación en que las varianzas de cada componente se multiplican por 2 y las covarianzas pasan de 0 a 0.5 y -0.5 respectivamente. En cada caso, se generan 5000 series. Entonces, se aplican los dos procedimientos con el valor crítico del 5 % de la Tabla 4.2. Los resultados se muestran en la Tabla 4.7, con el mismo diseño que antes. El caso con $\Omega = I$ muestra el error de tipo I de los procedimientos, que es alrededor del 5 % en todos los tamaños muestrales considerados. Cuando $T = 100$, el procedimiento cusum parece funcionar mejor que el LRT. Para $T = 200$ el procedimiento cusum es ligeramente superior al LRT, pero cuando $T = 500$, la frecuencia de detección de un punto de cambio es mayor que el 90 % y existe un pequeño incremento de la detección de dos cambios en el procedimiento cusum, lo que ya ocurría en el caso de cambios de varianzas. Las estimaciones del momento de cambio, h , son bastante parecidas en ambos procedimientos.

Finalmente, se estudia la potencia de los estadísticos cuando existe además un cambio en las matrices de parámetros, lo que se denominará un cambio estructural. Sea Y_t una serie generada por el siguiente modelo:

$$\begin{cases} \Phi_1(B) Y_t = C_1 + \Theta_1(B) A_t & t < h \\ \Phi_2(B) Y_t = C_2 + \Theta_2(B) (A_t + W S_t^{(h)} A_t) & t \geq h \end{cases},$$

tal que la matriz de covarianzas y los polinomios del modelo cambian en tiempo $t = h$. Los polinomios $\Phi_1(B)$, $\Phi_2(B)$, $\Theta_1(B)$ y $\Theta_2(B)$ se suponen que verifican las condiciones de estacionaridad e invertibilidad de la sección 4.2. Si los procedimientos tienen buenas propiedades de potencia para detectar cambios de la matriz de covarianzas bajo un cambio

Tabla 4.5: Resultados para el modelo 1 y dos cambios de varianza.

Procedimiento LRT												
Ω_1	Ω_2	T	h_1	h_2	frecuencia				\hat{h}_1		\hat{h}_2	
					0	1	2	≥ 3	Med.	Mad	Med.	Mad
$2 \times I$	I	100	33	66	75.2	18.2	6.6	0	33	4	68	4
		200	66	133	51.1	17.8	30.7	0.4	66	3	133	4
		500	166	333	6.6	3.6	85.0	4.8	166	4	333	4
$3 \times I$	I	100	33	66	33.9	14.8	50.9	0.4	33	2	66	2
		200	66	133	2.6	2.0	92.4	3.0	66	2	133	2
		500	166	333	0	0	82.6	17.4	166	1	333	1
$2 \times I$	$1/2 \times I$	100	33	66	9.4	67.5	22.8	0.2	32	3	66	1
		200	66	133	3.8	37.7	57.9	0.6	65	4	133	1
		500	166	333	17.4	2.4	76.4	3.8	166	4	333	1
$1/2 \times I$	$2 \times I$	100	33	66	12.4	86.0	1.6	0	21	4	66	1
		200	66	133	9.2	71.1	19.0	0.6	59	7	133	1
		500	166	333	21.4	9.0	59.9	9.6	163	5	334	1
$3 \times I$	$1/3 \times I$	100	33	66	3.6	23.4	72.1	0.8	33	2	66	0
		200	66	133	4.6	3.8	88.6	3.0	66	2	133	0
		500	166	333	0	1.0	81.2	17.8	166	1	333	0
$1/3 \times I$	$3 \times I$	100	33	66	3.8	83.0	12.4	0.8	26	4	66	1
		200	66	133	9.0	19.8	63.3	7.8	63	3	133	0
		500	166	333	0.8	1.6	78.6	19.0	164	2	333	0
Procedimiento Cusum												
Ω_1	Ω_2	n	h_1	h_2	frecuencia				\hat{h}_1		\hat{h}_2	
					0	1	2	≥ 3	Med.	Mad	Med.	Mad
$2 \times I$	I	100	33	66	75.4	7.4	17.2	0	35	2	64	2
		200	66	133	31.3	2.8	64.9	1.0	68	2	132	2
		500	166	333	0.6	0.2	91.8	7.4	167	3	331	3
$3 \times I$	I	100	33	66	27.5	3.2	68.9	0.4	34	1	65	1
		200	66	133	0.2	0	95.0	4.8	67	1	132	2
		500	166	333	0	0.2	91.4	8.4	167	1	331	2
$2 \times I$	$1/2 \times I$	100	33	66	20.8	31.5	47.5	0.2	35	3	65	1
		200	66	133	26.5	5.2	65.5	2.8	67	3	132	1
		500	166	333	23.4	0.6	70.5	5.4	168	4	332	1
$1/2 \times I$	$2 \times I$	100	33	66	12.6	72.7	14.0	0.6	31	2	67	1
		200	66	133	30.9	22.6	41.7	4.8	64	3	134	1
		500	166	333	23.4	3.6	65.3	7.6	165	3	334	1
$3 \times I$	$1/3 \times I$	100	33	66	16.2	4.4	78.4	1.0	34	1	65	1
		200	66	133	8.2	2.2	86.0	3.6	67	1	132	1
		500	166	333	0.2	0.8	92.0	7.0	167	2	332	1
$1/3 \times I$	$3 \times I$	100	33	66	18.4	43.9	34.5	3.2	32	1	67	1
		200	66	133	9.2	7.4	78.6	4.8	65	2	134	1
		500	166	333	0.8	4.0	87.8	7.4	165	2	334	1

Tabla 4.6: Resultados para el modelo 2 y dos cambios de varianza.

Procedimiento LRT												
Ω_1	Ω_2	T	h_1	h_2	frecuencia				\hat{h}_1		\hat{h}_2	
					0	1	2	≥ 3	Med.	Mad	Med.	Mad
$2 \times I$	I	100	33	66	67.9	21.8	10.2	0	32	4	67	3
		200	66	133	3.6	6.4	80.4	9.6	64	2	133	0
		500	166	333	1.2	0.2	93.8	4.8	166	4	333	3
$3 \times I$	I	100	33	66	19.4	13.2	66.5	0.8	33	2	66	1
		200	66	133	0.2	0.2	95.0	4.6	66	1	133	1
		500	166	333	0	0	73.3	26.7	166	1	333	1
$2 \times I$	$1/2 \times I$	100	33	66	2.8	65.1	32.1	0	32	3	66	1
		200	66	133	3.8	25.1	68.3	2.8	65	4	133	1
		500	166	333	7.6	1.0	85.8	5.6	166	3	333	1
$1/2 \times I$	$2 \times I$	100	33	66	3.0	93.0	4.0	0	24	6	66	0
		200	66	133	7.0	61.3	29.9	1.8	61	5	134	1
		500	166	333	10.2	3.0	71.1	15.6	163	4	334	1
$3 \times I$	$1/3 \times I$	100	33	66	3.6	13.2	82.4	0.8	33	1	66	0
		200	66	133	1.6	1.2	90.9	6.2	66	1	133	0
		500	166	333	0	0	78.8	21.2	166	1	333	0
$1/3 \times I$	$3 \times I$	100	33	66	4.8	76.4	18.4	0.4	29	4	66	0
		200	66	133	4.8	5.6	78.6	11.0	64	2	133	0
		500	166	333	0	0.4	81.0	18.6	165	1	333	0
Procedimiento Cusum												
Ω_1	Ω_2	T	h_1	h_2	frecuencia				\hat{h}_1		\hat{h}_2	
					0	1	2	≥ 3	Med.	Mad	Med.	Mad
$2 \times I$	I	100	33	66	54.9	9.8	34.9	0.4	34	2	65	1
		200	66	133	3.6	4.6	88.0	3.8	65	1	133	0
		500	166	333	0	0.2	92.6	7.2	167	2	332	2
$3 \times I$	I	100	33	66	6.4	0.8	90.6	2.2	34	1	65	1
		200	66	133	0.2	0	95.2	4.6	66	1	132	1
		500	166	333	0	0	92.0	8.0	167	1	333	1
$2 \times I$	$1/2 \times I$	100	33	66	15.0	18.8	65.7	0.4	34	1	65	1
		200	66	133	20.8	1.4	74.3	3.4	67	2	132	1
		500	166	333	10.0	0.4	81.4	8.2	167	2	332	1
$1/2 \times I$	$2 \times I$	100	33	66	14.4	59.5	23.8	2.2	32	2	66	0
		200	66	133	19.4	9.0	64.9	6.6	65	2	133	0
		500	166	333	12.8	3.0	78.8	5.4	165	2	334	1
$3 \times I$	$1/3 \times I$	100	33	66	13.4	0.8	84.4	1.4	33	1	66	0
		200	66	133	2.0	0.8	93.4	3.8	67	1	133	0
		500	166	333	0	0	90.6	9.4	167	1	333	0
$1/3 \times I$	$3 \times I$	100	33	66	12.4	24.6	58.7	4.2	33	1	66	0
		200	66	133	4.6	3.0	85.0	7.4	65	1	133	0
		500	166	333	0	0.8	90.8	8.4	165	1	333	0

Tabla 4.7: Resultados para el modelo 1 con cambio de varianzas y de correlaciones.

Procedimiento LRT								Procedimiento Cusum				
Ω	T	h	frecuencia			\hat{h}		frecuencia			\hat{h}	
			0	1	≥ 2	Med.	Mad	0	1	≥ 2	Med.	Mad
I	100	—	95.4	4.6	0	—	—	94.6	5.4	0	—	—
I	200	—	95.0	5.0	0	—	—	95.4	4.4	0.2	—	—
I	500	—	94.6	5.2	0.2	—	—	94.6	5.0	0.4	—	—
Ω_1	100	25	63.3	36.3	0.4	26	6	60.7	38.1	1.2	34	7
		50	45.9	53.3	0.8	50	5	18.0	81.2	0.8	52	3
		75	58.7	40.7	0.6	75	3	31.3	67.9	0.8	74	3
Ω_1	200	50	26.1	71.9	2.0	51	4	21.0	77.6	1.4	56	6
		100	9.2	89.0	1.8	101	5	1.6	95.8	2.6	103	4
		150	16.6	81.8	1.6	151	4	5.4	91.8	2.8	149	4
Ω_1	500	125	0.2	96.4	3.4	126	4	0.2	93.0	6.8	129	6
		250	0	97.4	2.6	251	3	0	94.0	6.0	252	4
		375	0	96.8	3.2	376	3	0	93.0	7.0	375	4
Ω_2	100	25	65.3	34.1	0.6	26	5	60.7	38.1	1.2	36	7
		50	47.1	52.3	0.6	51	6	20.8	78.6	0.6	52	3
		75	53.9	45.3	0.8	75	4	31.3	68.1	0.6	74	3
Ω_2	200	50	20.2	77.8	2.0	51	5	15.4	81.8	2.8	56	6
		100	9.4	87.8	2.8	101	4	0.4	97.6	2.0	102	4
		150	15.8	83.0	1.2	151	5	5.4	92.2	2.4	150	4
Ω_2	500	125	0.4	98.0	1.6	126	4	0.2	95.4	4.4	130	6
		250	0	97.2	2.8	251	3	0	94.8	5.2	252	4
		375	0	96.4	3.6	376	3	0.2	93.6	6.2	375	4

Tabla 4.8: Modelos para el estudio de simulación con cambios estructurales.

Π_1		Π_2		Ω_1		Ω_2	
0.6	0.2	0.3	0.4	2	0.5	2	-0.5
0.2	0.4	0.4	0.7	0.5	2	-0.5	2

Tabla 4.9: Resultados para cambios estructurales.

			Procedimiento LRT					Procedimiento Cusum				
Ω	T	h	frecuencia			\hat{h}		frecuencia			\hat{h}	
			0	1	≥ 2	Med.	Mad	0	1	≥ 2	Med.	Mad
Ω_1	100	25	74.1	25.3	0.6	24	6	73.3	26.5	0.2	41	10
		50	54.3	45.3	0.4	50	5	28.1	71.5	0.4	52	3
		75	54.1	45.7	0.2	75	4	30.5	69.1	0.4	74	3
Ω_1	200	50	43.7	54.9	1.4	50	8	43.9	54.5	1.6	63	12
		100	15.8	81.8	2.4	101	8	5.0	92.6	2.4	102	5
		150	14.2	80.4	5.4	152	5	5.4	91.6	3.0	149	4
Ω_1	500	125	3.2	91.2	5.6	126	6	2.0	94.0	4.0	132	11
		250	0	91.6	8.4	251	4	0	93.2	6.8	252	5
		375	0.2	72.5	27.3	377	4	0	94.0	6.0	375	5
Ω_2	100	25	73.7	26.1	0.2	25	5	75.8	23.8	0.4	38	10
		50	53.9	44.9	1.2	49	8	31.5	67.7	0.8	52	4
		75	48.7	50.5	0.8	75	4	31.3	67.9	0.8	74	4
Ω_2	200	50	44.5	53.1	2.4	50	8	40.9	58.3	0.8	62	12
		100	12.4	85.6	2.0	101	7	3.0	93.0	4.0	102	5
		150	11.2	83.4	5.4	151	4	5.0	92.8	2.2	149	5
Ω_2	500	125	5.8	91.4	2.8	126	8	3.2	93.4	3.4	132	10
		250	0	90.2	9.8	251	5	0	96.0	4.0	253	5
		375	0.2	71.9	27.9	377	4	0	92.0	8.0	375	4

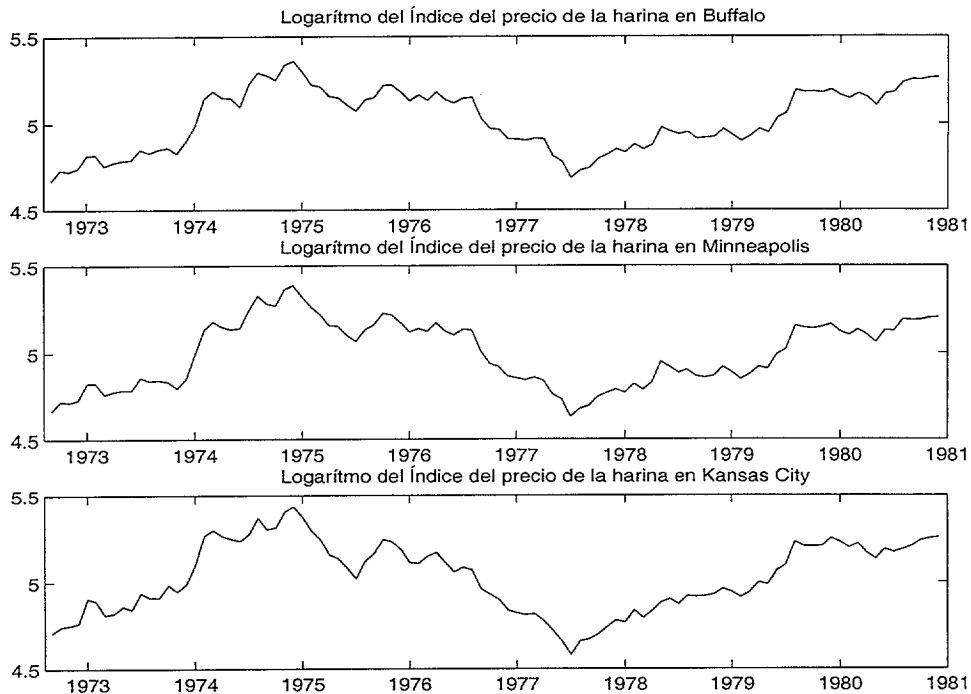


Figura 4.2: Índices de precios de la harina en tres ciudades de U.S.A.

estructural, éstos pueden ser un arma útil para detectar además cambios estructurales. Se consideran los modelos en la Tabla 4.8 y los tamaños muestrales $T = 100, 200$ y 500 , para $k = 2$. Los cambios se introducen transformando la matriz de covarianzas original, $\Sigma = I$, en Ω_1 y Ω_2 , y los polinomios autorregresivos Π_1 en Π_2 . Para cada caso, se generan 5000 series y se aplican los procedimientos con los valores críticos al 5% de la Tabla 4.2. Los resultados aparecen en la Tabla 4.9. Se concluye que ambos procedimientos tienen un pequeño decrecimiento en la potencia para pequeñas muestras, especialmente en el caso en que $h = [0.25T]$, pero no pierden potencia para tamaños muestrales grandes, aquí $T = 500$.

4.6. Ejemplos

4.6.1. El precio de la harina

Se considera la serie bivalente de los logaritmos de los índices mensuales de precios de la harina en tres ciudades de Estados Unidos en el periodo de Agosto de 1972 a Noviembre de 1980. Esta serie vectorial fue analizada por Tiao y Tsay (1989), Grubb (1992) y Lütkepohl y Poskitt (1996) y aparece en la Figura 4.2. Tiao y Tsay (1989) ajustan un vector ARMA(1,1) con restricciones a la serie mientras que Grubb (1992), por el criterio de información de

Tabla 4.10: Matriz de parametros AR ($\hat{\Pi}$), matriz de covarianza estimada ($\hat{\Sigma}$) y matriz de correlaciones estimada (R) para los tres modelos ajustados para la serie del precio de la harina. Errores estándar de los coeficientes aparecen bajo paréntesis.

Mod.	VAR(1)			Cambio en $t = 33$ Abril, 1975			Pr. parte (1-32)			Seg. Parte (33-100)		
$\hat{\Pi}$	-0.86	1.01	0	-0.83	0.96	0	-0.61	0.95	0	-0.28	0.26	0
	(0.17)	(0.18)		(0.15)	(0.16)		(0.14)	(0.20)		(0.11)	(0.11)	
	-0.43	0.62	0	-0.48	0.64	0	0	0.40	0	0	0	0
	(0.17)	(0.19)		(0.13)	(0.14)			(0.16)				
	0	0.25	0	0	0.21	0	0	0.43	0	0	0	0
		(0.10)			(0.11)			(0.18)				
$10^2 \hat{\Sigma}$	0.20	0.21	0.20	0.20	0.21	0.20	0.24	0.26	0.27	0.19	0.20	0.17
	0.21	0.24	0.22	0.21	0.23	0.22	0.26	0.30	0.31	0.20	0.21	0.19
	0.20	0.22	0.27	0.20	0.22	0.26	0.27	0.31	0.37	0.17	0.19	0.21
R	1	0.95	0.86	1	0.97	0.87	1	0.96	0.90	1	0.99	0.85
	0.95	1	0.86	0.97	1	0.89	0.96	1	0.93	0.99	1	0.90
	0.86	0.86	1	0.87	0.89	1	0.90	0.93	1	0.85	0.90	1

Tabla 4.11: Resumen de los procedimientos LRT y cusum para los datos de la harina.

Método	VAR(1)	LRT W diag.	Cusum W diag.	LRT W triang.			Cusum W triang.		
h	—	28	60	33			33		
$\Lambda_{\max}/\Gamma_{\max}$	—	13.26	0.63	28.95			1.78		
\widehat{W}	—	—	—	-0.14	0	0	-0.14	0	0
				0.40	-0.53	0	0.40	-0.53	0
				-0.02	-0.17	-0.09	-0.02	-0.17	-0.09
AIC	-14.01	-14.01	-14.01	-27.39			-27.39		
BIC	-13.77	-13.77	-13.77	-26.99			-26.99		

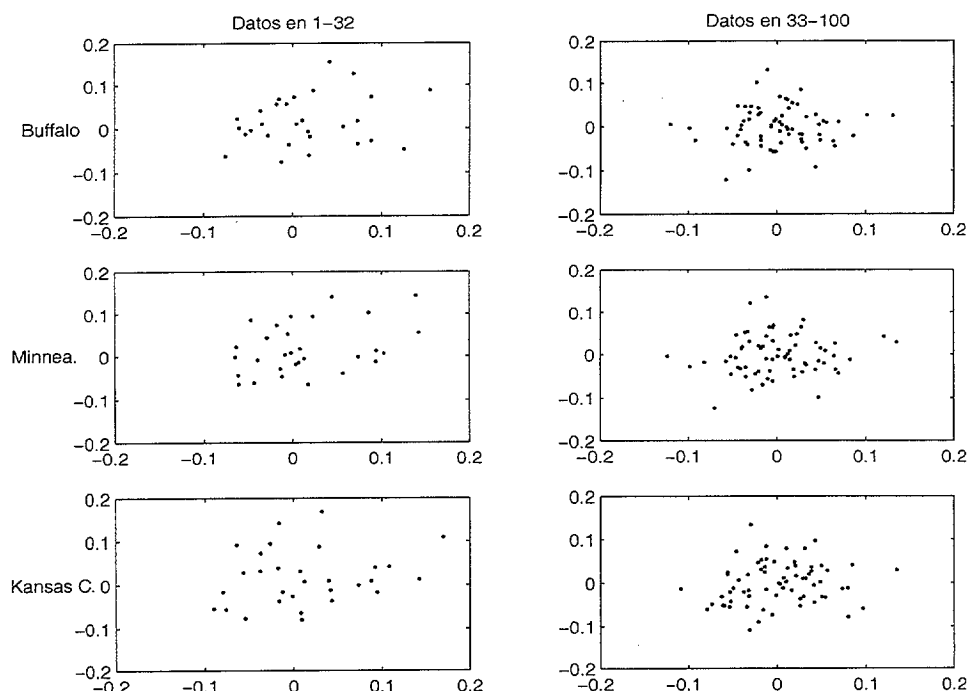


Figura 4.3: Gráfico de puntos de la primera diferencia del logaritmo de cada Índice frente al primer retardo de la misma serie.

Akaike (AIC), selecciona un vector $AR(2)$ con restricciones. Lütkepohl y Poskitt (1996) investigan cointegración en estas series mediante el test de Johansen en un modelo vector $AR(2)$, rechazando la hipótesis nula de cointegración. Entonces, ajustan un vector $AR(1)$ para la serie diferenciada, que se muestra en la segunda columna de la Tabla 4.10.

Utilizando este modelo, se aplican los procedimientos LRT y cusum para cambios de varianza a los datos. La Tabla 4.11 resume los resultados. Las columnas 3 y 4 muestran que los máximos de los estadísticos LRT y cusum no son significativos, por lo que no se detecta ningún cambio de varianza. Entonces se aplican los procedimientos para el caso de cambio de varianzas y de correlaciones para los datos y ambos procedimientos detectan un cambio en $t = 33$ (Abril, 1975) como se muestra en las columnas 5 y 6 de la Tabla 4.11. La estimación de W se obtiene como en (4.15). Teniendo en cuenta que no existe evidencia de cambios de varianza, aparentemente el cambio acontece en la estructura de correlación entre los componentes de la serie. En la Tabla 4.11 se incluyen los valores de los criterios de información Akaike y Bayesiano (AIC y BIC) para cada modelo. El valor de ambos criterios se reduce cuando el cambio en la matriz de covarianzas se incluye en la estimación, y los dos criterios indican que el modelo con un cambio de varianza en $t = 33$ (Abril, 1975) parece ser más apropiado para los datos. El modelo final estimado se muestra en la tercera columna de la Tabla 4.10.

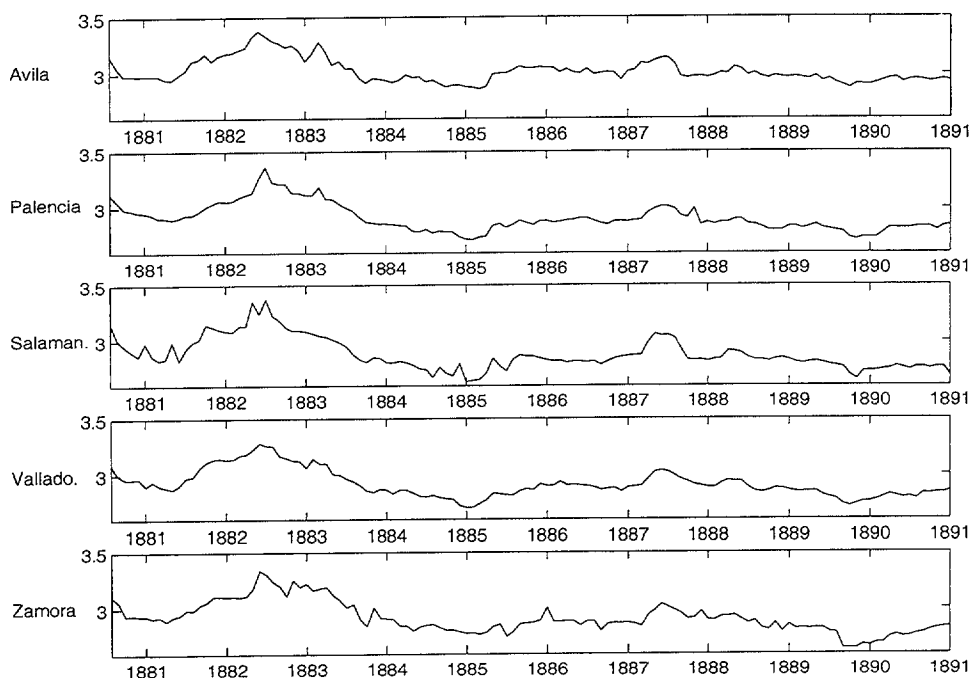


Figura 4.4: Índices de precios del trigo en cinco provincias de Castilla, España

Finalmente, se estima un modelo vector AR(1) a las submuestras 1-32 y 33-100. Los dos modelos estimados con sus errores estándar se muestran en la cuarta y quinta columna de la Tabla 4.10. Como puede verse, los parámetros del modelo y las matrices de covarianzas residuales son diferentes en ambos modelos por lo que se concluye que la serie tiene un cambio estructural en $t = 33$ (Abril, 1975). El modelo para la segunda parte de la serie es aparentemente un paseo aleatorio y la matriz de covarianzas residual tiene valores menores que la obtenida en la primera parte de la serie. Para entender mejor el cambio, las tres filas de la figura 4.3 muestran los gráficos de puntos de la diferencia del logaritmo del Índice en cada una de las ciudades frente al primer retardo de la misma serie. En cada columna aparece una de las submuestras, 1-32 y 33-100. Se puede comprobar que la primera columna muestra puntos con una pequeña correlación positiva, mientras que en la segunda columna los datos aparentemente no muestran correlación.

4.6.2. El precio del trigo

Se considera la serie en logaritmos del índice mensual de precios del trigo de cinco provincias en Castilla en España, en el periodo de Julio de 1880 a Diciembre de 1890. Esta serie vectorial fue analizada en Peña y Box (1987) y se muestra en la Figura 4.3. En primer lugar

Tabla 4.12: Matriz de parametros AR ($\hat{\Pi}$), matriz de covarianza estimada ($\hat{\Sigma}$) y matriz de correlaciones estimada (R) para los tres modelos ajustados para la serie del precio del trigo. Errores estándar de los coeficientes aparecen bajo paréntesis.

Modelo	VAR(1)					3 C. V.				
$\hat{\Pi}_1$	0.87 (0.07)	0	0	0.44 (0.12)	0	0.92 (0.05)	-0.34 (0.10)	0	0.48 (0.09)	0
	0.13 (0.05)	0.32 (0.07)	0	0.43 (0.09)	0	0.10 (0.03)	0.39 (0.05)	0.06 (0.03)	0.37 (0.05)	0
	0	-0.26 (0.13)	0.40 (0.09)	0.82 (0.16)	0	0	-0.73 (0.18)	0.41 (0.09)	1.10 (0.16)	0
	0.17 (0.05)	-0.21 (0.08)	0.15 (0.06)	0.85 (0.10)	0	0.17 (0.05)	-0.50 (0.10)	0.14 (0.05)	1.05 (0.09)	0
	0	0	0	0.46 (0.14)	0.47 (0.08)	0	0	0.14 (0.06)	0.55 (0.10)	0.54 (0.11)
$10^3 \hat{\Sigma}$	1.44	0.46	0.59	0.68	0.59	0.64	0.21	0.34	0.30	0.35
	0.46	0.84	0.48	0.32	0.40	0.21	0.18	0.31	0.24	0.26
	0.59	0.48	2.39	0.59	0.62	0.34	0.31	1.94	0.32	0.71
	0.68	0.32	0.59	0.91	0.50	0.30	0.24	0.32	0.61	0.37
	0.59	0.40	0.62	0.50	2.00	0.35	0.26	0.71	0.37	0.77
R	1	0.41	0.31	0.59	0.34	1	0.61	0.30	0.48	0.49
	0.41	1	0.33	0.36	0.30	0.61	1	0.52	0.72	0.69
	0.31	0.33	1	0.40	0.28	0.30	0.52	1	0.29	0.58
	0.59	0.36	0.40	1	0.37	0.48	0.72	0.29	1	0.53
	0.34	0.30	0.28	0.37	1	0.49	0.69	0.58	0.53	1

Tabla 4.13: Resumen de los procedimientos LRT y cusum para cambios de varianzas para los datos del trigo.

Método	VAR(1)	LRT			Cusum	
h	—	23	90	22	41	90
$\Lambda_{\max}/\Gamma_{\max}$	—	30.99	90.93	1.61	1.71	1.70
$\widehat{W}_{I.C.}(1,1)$	—	0	-0.48 (-0.57,-0.37)	0.54 (0.02,1.22)	-0.38 (-0.60,-0.13)	-0.38 (-0.52,-0.21)
$\widehat{W}_{I.C.}(2,2)$	—	0.71 (-0.46,-0.21)	-0.34 (-0.46,-0.21)	1.16 (0.43,2.10)	0	-0.33 (-0.49,-0.15)
$\widehat{W}_{I.C.}(3,3)$	—	-0.30 (-0.52,-0.04)	-0.58 (-0.66,-0.50)	0	0	-0.53 (-0.64,-0.40)
$\widehat{W}_{I.C.}(4,4)$	—	0	-0.29 (-0.42,-0.15)	0	-0.37 (-0.59,-0.11)	0
$\widehat{W}_{I.C.}(5,5)$	—	0.51 (0.03,1.05)	-0.29 (-0.42,-0.14)	1.12 (0.40,2.04)	-0.43 (-0.63,-0.20)	0
AIC	-19.84	-32.85			-37.52	
BIC	-19.52	-32.28			-36.98	

se investiga si existen relaciones de cointegración en las series utilizando el test de Johansen. Utilizando el BIC se selecciona un modelo vector AR(1) con tres relaciones de cointegración, lo que coincide con los dos factores encontrados por Peña y Box (1987). Entonces, se aplican los procedimientos para cambios de varianza a los datos bajo un modelo vector AR(1) sin restricciones y después en un modelo de corrección del error con tres relaciones de cointegración. Como los resultados obtenidos con ambos modelos son similares, sólo mostramos los obtenidos con el vector AR(1) sin restricciones y que aparecen en la segunda fila de la Tabla 4.12. La Tabla 4.13 resume los resultados. El procedimiento LRT detecta un cambio de varianzas en $h = 22$ (Abril, 1882) y otro en $h = 90$ (Diciembre, 1888). El procedimiento cusum detecta tres cambios en $h = 22$ (Abril, 1882), $h = 41$ (Noviembre, 1883) y $h = 90$ (Diciembre, 1888). La estimación de los cambios y sus intervalos de confianza aparecen en la Tabla 4.13. Si algún cambio en alguna componente no es significativo, se representa por un 0. Los valores mínimos de los criterios AIC y BIC se obtienen para el modelo propuesto por el procedimiento cusum con tres cambios.

A continuación, se aplican los procedimientos LRT y cusum para cambios de varianzas y de correlaciones que son resumidos en la Tabla 4.14. El procedimiento LRT detecta un cambio en $h = 88$ (Octubre, 1888), mientras que el procedimiento cusum detecta tres cambios similares a los anteriores en $h = 22$ (Abril, 1882), $h = 35$ (Mayo, 1883) y $h = 90$ (Diciembre, 1888). La estimación de los cambios aparecen en la Tabla 4.14. Los mínimos de los valores de los criterios AIC y BIC corresponden a los modelos propuestos por el algoritmo cusum con tres cambios.

Tabla 4.14: Resumen de los procedimientos LRT y cusum para cambios de varianzas y de correlaciones para los datos de la harina.

Método	h	$\Lambda_{\max}/\Gamma_{\max}$	\widehat{W}					AIC	BIC
VAR(2)	—	—	—					-19.84	-19.52
LRT	88	82.79	-0.50	0	0	0	0	-29.33	-27.36
			-0.17	0.09	0	0	0		
			0.07	-0.42	-0.53	0	0		
			-0.07	0.00	-0.03	-0.20	0		
			-0.20	-0.33	-0.16	0.72	-0.32		
Cusum	22	2.35	0.74	0	0	0	0	-38.85	-37.43
			-0.51	2.00	0	0	0		
			0.53	-1.67	0.31	0	0		
			0.81	-2.55	0.35	0.50	0		
			0.22	-2.35	-0.33	0.06	1.48		
	35	2.26	-0.41	0	0	0	0		
			0.07	-0.42	0	0	0		
			0.11	0.32	-0.33	0	0		
			-0.12	0.30	-0.05	-0.41	0		
			0.00	0.25	0.19	-0.06	-0.28		
	90	2.03	-0.39	0	0	0	0		
			0.03	-0.32	0	0	0		
			-0.08	-0.21	-0.40	0	0		
			-0.04	0.05	-0.06	0.00	0		
			-0.08	-0.02	-0.33	0.61	-0.32		

El modelo final seleccionado por el BIC (aunque el AIC da el mismo resultado), es el obtenido por el procedimiento cusum que contrasta cambios en las varianzas y en las correlaciones. Los parámetros estimados para este modelo vector AR(1) aparecen en la tercera fila de la Tabla 4.12. Para entender mejor los cambios, las cinco filas de la figura 4.5 muestran los gráficos de puntos de la diferencia del logaritmo del Índice en cada una de las ciudades frente al primer retardo de la misma serie. En cada columna aparece una de las submuestras, 1-21, 22-34, 35-89 y 90-126. Las diferencias entre las submuestras parecen claras.

4.7. Conclusiones

En este capítulo se estudian cambios en la matriz de covarianzas marginal de una serie temporal multivariante que sigue un modelo ARMA vectorial. Se analizan las dos propuestas más conocidas en el caso univariante: el procedimiento basado en el contraste de razón de

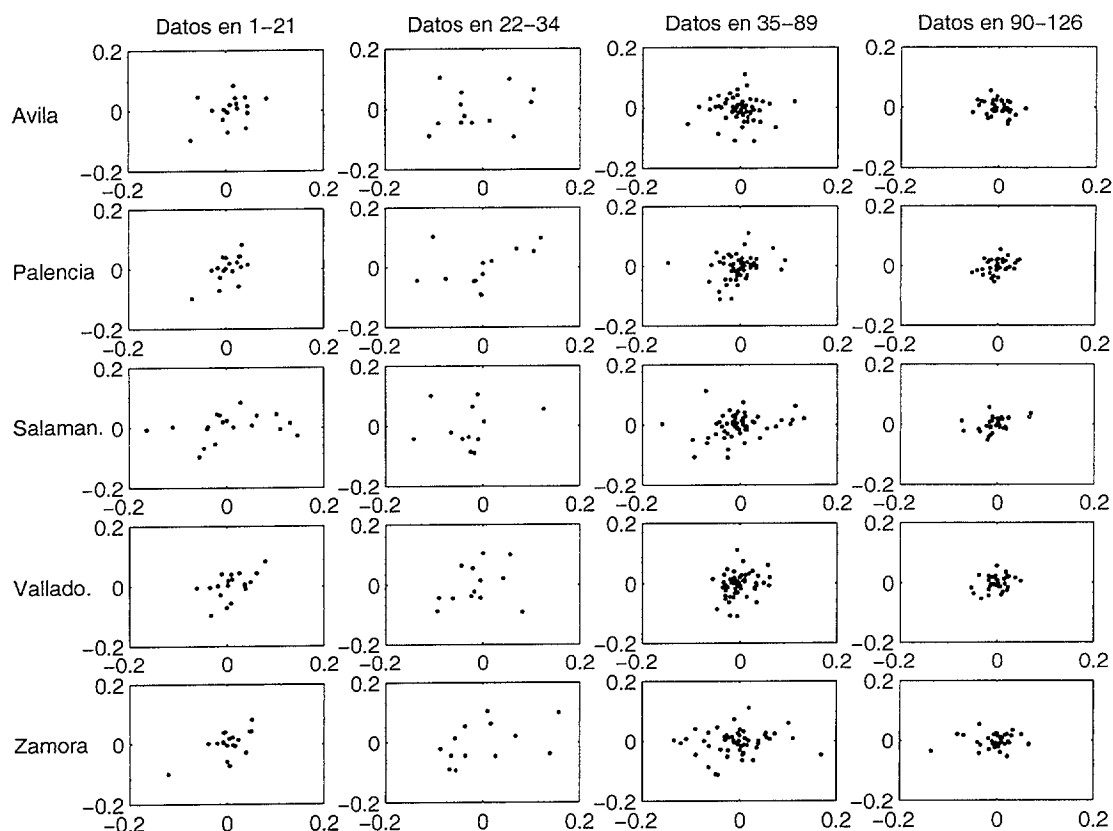


Figura 4.5: Gráfico de puntos de la primera diferencia del logaritmo de cada Índice frente al primer retardo de la misma serie.

verosimilitudes propuesto por Tsay (1988) y el procedimiento de segmentación basado en un estadístico cusum propuesto por Inclán y Tiao (1994). Para ello, se propone el uso de dos estadísticos multivariantes para detectar cambios en las varianzas de las componentes de la serie multivariante y otros dos estadísticos que además contrastan la hipótesis de cambios en las varianzas y correlaciones de las componentes del vector de series. Se proponen sendos algoritmos para cada tipo de estadístico. Un estudio de simulación compara ambos algoritmos y se demuestra que en general, el algoritmo de segmentación para estadísticos cusum es superior al algoritmo basado en el contraste de razón de verosimilitudes. El funcionamiento práctico de ambos métodos es ilustrado en dos series previamente analizadas en la literatura de precios de la harina y del trigo.

4.8. Apéndice

Demostración del Lema 4.1. El desarrollo de Taylor de segundo orden del radio D_m/D_T en el valor $(E[D_m], E[D_T])$ es:

$$E\left[\frac{D_m}{D_T}\right] = \frac{E[D_m]}{E[D_T]} - \frac{E[D_m D_T]}{E[D_T]^2} + \frac{E[D_m] E[D_T^2]}{E[D_T]^3} + o(T^{-1}).$$

Teniendo en cuenta que $\text{tr}(\Lambda) = k$, como $E[D_m] = mk$, y

$$\begin{aligned} E[D_m D_T] &= E\left[\left(\sum_{t=1}^m C'_t C_t\right) \left(\sum_{l=1}^T C'_l C_l\right)\right] = \\ &= \sum_{t=1}^m \sum_{l=1}^T E[(C'_t C_t)(C'_l C_l)] = m[2\text{tr}(\Lambda^2) + k^2], \end{aligned}$$

el radio $E[D_m/D_T]$ se puede escribir como:

$$E\left[\frac{D_m}{D_T}\right] = \frac{m}{T} - \frac{m[2\text{tr}(\Lambda^2) + k^2]}{T^2 k^2} + \frac{T[2\text{tr}(\Lambda^2) + k^2]mk}{T^3 k^3} + o(T^{-1}) = \frac{m}{T} + o(T^{-1}),$$

y por lo tanto, $E[F_m] = o(T^{-1})$. ■

Demostración del Teorema 4.2. Sea $\xi_m = C'_m C_m - k$, tal que $E[\xi_m] = 0$, y,

$$\sigma^2 = E[\xi_m^2] = E[(C'_m C_m)^2] - k^2 = 2\text{tra}(\Lambda^2).$$

Sea $X_T(r) = \frac{1}{\sigma\sqrt{T}}S_{[Tr]} + (Tr - [Tr])\frac{1}{\sigma\sqrt{T}}\xi_{[Tr]+1}$, donde $S_T = \sum_{i=1}^T \xi_i$. Por el Teorema de Donsker, $X_T \xrightarrow{D} M$, por lo que $\{X_T(r) - rX_T(1)\} \xrightarrow{D} M_0$, ver Billingsley (1968, Th. 10.1 and Th. 5.1). Sea $Tr = m$, $m = 1, \dots, T$. Entonces,

$$X_T(r) - rX_T(1) = \frac{1}{\sigma\sqrt{T}}S_{[Tr]} + (Tr - [Tr])\frac{1}{\sigma\sqrt{T}}\xi_{[Tr]+1} - r\frac{1}{\sigma\sqrt{T}}S_T = \frac{1}{\sigma\sqrt{T}}\left(S_m - \frac{m}{T}S_T\right).$$

Como $S_T = \sum_{t=1}^T C'_t C_t - Tk$ y $S_m = \sum_{t=1}^m C'_t C_t - mk$, se obtiene,

$$S_m - \frac{m}{T}S_T = \left(\sum_{t=1}^T C'_t C_t\right) F_m.$$

Entonces, cuando $T \rightarrow \infty$,

$$\frac{1}{\sigma\sqrt{T}}\left(\sum_{t=1}^T C'_t C_t\right) F_m \xrightarrow{D} M^0.$$

Por lo tanto, como $\frac{1}{T} \sum_{t=1}^T C'_t C_t \rightarrow k$,

$$\frac{1}{\sigma \sqrt{T}} \left(\sum_{t=1}^T C'_t C_t \right) F_m = \sqrt{\frac{T}{2}} \frac{k}{\sqrt{\text{tr}(\Lambda^2)}} \frac{\frac{1}{T} \sum_{t=1}^T C'_t C_t}{k} F_m \xrightarrow{D} M^0$$

que prueba el resultado establecido. ■

Capítulo 5

Criterios de selección de modelos y discriminación cuadrática en series temporales

5.1. Introducción

La mayoría de los criterios de selección de modelos en series temporales pueden ser escritos como,

$$\min_r \{ \log \hat{\sigma}_r^2 + r \times C(T, r) \}, \quad (5.1)$$

donde $\hat{\sigma}_r^2$ es el estimador máximo verosímil de la varianza residual, r es el número de parámetros estimados para la función de medias del proceso, T es el tamaño muestral y $C(T, r)$ es una función que converge a 0 cuando $T \rightarrow \infty$. Como ya mencionamos en el capítulo 2, estos criterios se pueden clasificar en dos grupos. El primero incluye a los criterios consistentes que, bajo la hipótesis de que los datos han sido generados por un proceso ARMA de ordenes finitos, tienen probabilidad de seleccionar el verdadero modelo que tiende a uno cuando el tamaño muestral crece. El criterio de información Bayesiano, BIC, de Schwarz (1978), donde $C(T, r) = \log(T)/T$, y el criterio de Hannan y Quinn (1979), HQC, donde $C(T, r) = 2m \log \log(T)/T$ con $m > 1$, son criterios consistentes. El segundo grupo incluye a los criterios eficientes, que asintóticamente seleccionan el orden que produce el menor error cuadrático de predicción. El criterio del error de predicción final, FPE, de Akaike (1969), donde $C(T, r) = r^{-1} \log(\frac{T+r}{T-r})$, el criterio de información de Akaike, AIC, de Akaike (1973), donde $C(T, r) = 2/T$ y el criterio de información de Akaike corregido, AICc, de Hurvich y Tsai (1989), donde $C(T, r) = \frac{1}{r} \frac{2(r+1)}{T-(r+2)}$, son criterios eficientes.

Como se comentó en el capítulo 2, estos criterios han sido derivados desde diferentes puntos de vista. En este capítulo se considera el problema de selección de modelos como un problema de discriminación y se demuestra que los criterios BIC, AIC y AICc pueden ser

derivados como aproximaciones de una regla de discriminación cuadrática. Este planteamiento además proporciona un término de corrección que mejora el comportamiento de estos criterios en muestras finitas, manteniendo sus propiedades asintóticas.

Una extensión no lineal de los modelos ARMA son los modelos autorregresivos por umbrales (TAR), ver Tong (1990). Estos modelos pueden explicar características interesantes que se encuentran en los datos reales, como ciclos asimétricos, fenómenos de salto, caos y otros. Una serie que sigue un modelo TAR es lineal en regímenes separados, por lo que la corrección considerada para modelos lineales puede ser fácilmente extendida a estos modelos no lineales. Criterios de selección para modelos TAR han sido estudiados en Wong y Li (1998), Kapetanios (2001) y De Gooijer (2001). Aquí se analiza el término de corrección obtenido por el enfoque propuesto y se demuestra que este término hace mejorar el comportamiento en muestras finitas de los criterios previos.

El resto de este capítulo está organizado como sigue. En la sección 5.2 se introduce la familia de criterios para modelos ARMA. La sección 5.3 discute el cálculo del término de corrección. La sección 5.4 incluye el término de corrección en los modelos TAR. Por último, en la sección 5.5 se explora el comportamiento de los criterios propuestos en un experimento de Monte Carlo y se demuestra que dichos criterios tienen mejor comportamiento que los criterios clásicos para todos los modelos considerados.

5.2. Una familia de criterios de selección de modelos basados en la regla de discriminación cuadrática

El problema de discriminación en series temporales se plantea como sigue. Se supone conocido que una serie temporal, $x = (x_1, \dots, x_T)'$, ha sido generada por uno de los modelos M_j , $j = 1, \dots, j_{\max}$. Desde un punto de vista Bayesiano también se tienen probabilidades a priori del modelo, $p(M_j)$. El objetivo es seleccionar el modelo generador de los datos dada la serie temporal. En el caso general, se supone que los modelos M_j son procesos Gaussianos dados por $x_t = \mu_{jt} + n_{jt}$, donde μ_{jt} son funciones de media deterministas y n_{jt} son modelos ARMA de media cero de la forma $\phi_j(B) n_{jt} = \theta_j(B) a_{jt}$, donde $\phi_j(B)$ y $\theta_j(B)$ son polinomios en el operador retardo B tal que $Bx_t = x_{t-1}$, sin raíces comunes. Los modelos se suponen casuales e invertibles. Las series a_{jt} son residuos ruido blanco con varianza σ_{ja}^2 . El problema de discriminación más simple es suponer que las funciones deterministas μ_{jt} son diferentes, pero las matrices de covarianzas Σ_j son todas iguales a Σ . Este caso corresponde a la situación en que todos los modelos tienen la misma estructura ARMA. Llamando $\mu_j = (\mu_{j1}, \dots, \mu_{jT})'$, esto es equivalente a considerar la hipótesis $M_j : x \in N_T(\mu_j, \Sigma)$, y se tiene que,

$$p(x | M_j) = (2\pi)^{-\frac{T}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (x - \mu_j)' \Sigma^{-1} (x - \mu_j) \right), \quad j = 1, \dots, j_{\max}.$$

Maximizar la verosimilitud de los datos implica minimizar la distancia de Mahalanobis entre los datos y el vector de medias marginal. La misma conclusión se obtiene desde un punto de vista Bayesiano suponiendo probabilidades a priori $p(M_j) = 1/j_{\max}$ iguales y maximizando la probabilidad a posteriori de seleccionar el verdadero modelo.

Aparece un caso más interesante cuando los modelos ARMA son diferentes, esto es, $M_j : x \in N_T(\mu_j, \Sigma_j)$, para $j = 1, \dots, j_{\max}$, donde Σ_j son las matrices de covarianza de x bajo cada modelo ARMA, n_{jt} . Entonces,

$$p(x | M_j) = (2\pi)^{-\frac{T}{2}} |\Sigma_j|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (x - \mu_j)' \Sigma_j^{-1} (x - \mu_j) \right)$$

la regla de clasificación cuadrática selecciona el modelo i si,

$$i = \arg \max_{1 \leq j \leq j_{\max}} (2\pi)^{-\frac{T}{2}} |\Sigma_j|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (x - \mu_j)' \Sigma_j^{-1} (x - \mu_j) \right) \quad (5.2)$$

y la regla Bayesiana selecciona el modelo i si,

$$i = \arg \max_{1 \leq j \leq j_{\max}} p(M_j) (2\pi)^{-\frac{T}{2}} |\Sigma_j|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (x - \mu_j)' \Sigma_j^{-1} (x - \mu_j) \right). \quad (5.3)$$

Estas reglas (5.2) y (5.3) se aproximan en esta sección para obtener versiones corregidas de los criterios AIC, AICc y BIC. Para ello, se considera el caso en que la serie temporal, $x = (x_1, \dots, x_T)'$, ha sido generada por la clase de procesos Gaussianos ARMA dados por,

$$x_t - \phi_1 x_{t-1} - \dots - \phi_p x_{t-p} = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}, \quad t = \dots, -1, 0, 1, \dots \quad (5.4)$$

donde a_t es una secuencia de variables aleatorias Gaussianas independientes de media cero y varianza $\sigma_{p,q}^2$. Se supone que $p \in \{0, \dots, p_{\max}\}$ y $q \in \{0, \dots, q_{\max}\}$, donde p_{\max} y q_{\max} son cotas máximas fijas. Se denota por $M_{p,q}$ al modelo ARMA(p, q) con vector de parámetros $\beta_{p,q} = (\phi_{1p}, \dots, \phi_{pp}, 0, \dots, 0, \theta_{1q}, \dots, \theta_{qq}, 0, \dots, 0)'$ y se define $\alpha_{p,q} = (\beta_{p,q}, \sigma_{p,q}^2)$, de tamaño $(p_{\max} + q_{\max} + 1) \times 1$. Se denotan a los parámetros del modelo que han generado los datos por $\alpha_0 = (\beta_0, \sigma_0^2)$. Sea $\hat{\beta}_{p,q}$ el estimador máximo verosímil del vector de parámetros $\beta_{p,q}$ suponiendo que el verdadero modelo es $M_{p,q}$ y sea $\hat{\sigma}_{p,q}^2$ el estimador máximo verosímil de la varianza de las innovaciones. La matriz de covarianzas de x suponiendo el modelo $M_{p,q}$ se puede escribir como $\Sigma_T(\alpha_{p,q}) = \sigma_{p,q}^2 Q_T(\beta_{p,q})$, donde $Q_T(\beta_{p,q})$ es una matriz $T \times T$ que depende de los parámetros $\beta_{p,q}$. Sea $Q_T(\beta_{p,q}) = L(\beta_{p,q}) L'(\beta_{p,q})$ la descomposición de Cholesky de $Q_T(\beta_{p,q})$. Se denotan, $a(\beta_{p,q}) = L(\beta_{p,q})^{-1} x$ y $S_x(\beta_{p,q}) = a(\beta_{p,q})' a(\beta_{p,q})$. Por último, se considera la siguiente hipótesis:

Hipótesis 1: Los modelos $M_{p,q}$ son casuales, invertibles y estacionarios y con polinomios $1 - \phi_1 B - \dots - \phi_p B^p$ y $1 - \theta_1 B - \dots - \theta_q B^q$ sin raíces comunes.

5.2.1. Una aproximación máximo verosímil

La regla discriminante (5.2) asigna los datos $x = (x_1, \dots, x_T)'$, al modelo $M_{p,q}$ con parámetros $\alpha_{p,q}$ que maximizan $p(x | M_{p,q}) = p(x | \alpha_{p,q})$. En la práctica, los parámetros son desconocidos y la primera idea es substituir los parámetros desconocidos $\alpha_{p,q}$ por sus estimadores máximo verosímiles, $\hat{\alpha}_{p,q}$. Es bien conocido que esta solución siempre selecciona el modelo con el mayor número de parámetros y, para evitarlo, se calculan los estimadores máximo verosímiles $\hat{\alpha}_{p,q}$ basados en los datos, x , para cada posible modelo y se selecciona aquel que maximiza,

$$E_{\alpha_0}[\log p(y|\hat{\alpha}_{p,q})] = \int p(y|\alpha_0) \log p(y|\hat{\alpha}_{p,q}) dy, \quad (5.5)$$

esto es, el modelo que maximiza la esperanza con respecto a las futuras observaciones generadas por el modelo verdadero. El modelo seleccionado por este criterio es aquel que minimiza la distancia de Kullback-Leibler entre el modelo estimado y el verdadero. En efecto, como,

$$E_{\alpha_0} \left[\log \frac{p(y|\alpha_0)}{p(y|\hat{\alpha}_{p,q})} \right] = \int p(y|\alpha_0) \log \frac{p(y|\alpha_0)}{p(y|\hat{\alpha}_{p,q})} dy \geq 0$$

esta integral es siempre positiva. Como $\int p(y|\alpha_0) \log p(y|\alpha_0) dy$ es desconocida pero constante, minimizar esta divergencia implica maximizar $\int p(y|\alpha_0) \log p(y|\hat{\alpha}_{p,q}) dy$ que es el criterio (5.5). Minimizar la distancia implica hacer $p(y|\hat{\alpha}_{p,q})$ lo más cercano posible a $p(y|\alpha_0)$. El criterio será calcular el logaritmo de la verosimilitud de cada modelo utilizando los estimadores $\hat{\alpha}_{p,q}$ y la esperanza con respecto a las futuras observaciones. El modelo seleccionado será aquel que obtiene la mayor esperanza con respecto a las observaciones futuras.

Este criterio está relacionado con el propuesto por Akaike (1973). Este autor propuso seleccionar la estimación que proporcione el mayor valor de,

$$E_{\hat{\alpha}_{p,q}} [E_y [\log p(y|\hat{\alpha}_{p,q})]] = \int \int p(y|\alpha_0) \log p(y|\hat{\alpha}_{p,q}) dy d\hat{\alpha}_{p,q}$$

donde $\hat{\alpha}_{p,q}$ e y se suponen independientes. Así, Akaike calcula el valor esperado con respecto a la distribución de las observaciones futuras y respecto a la distribución del estimador. El criterio aquí propuesto, es más sencillo y en la práctica lleva al mismo resultado, como se verá en el siguiente Lema cuya demostración se presenta en el apéndice.

Lema 5.1 *Bajo la hipótesis 1, la esperanza con respecto a la verdadera distribución de los datos del logaritmo de la probabilidad de x dados los parámetros está dada por:*

1. si los parámetros son evaluados en $\hat{\beta}_{p,q}$ y $\frac{T}{T-(p+q)}\hat{\sigma}_{p,q}^2$:

$$-\frac{T}{2} (\log 2\pi + 1) - \frac{T}{2} \log \hat{\sigma}_{p,q}^2 - \frac{1}{2} \log \left| Q_T \left(\hat{\beta}_{p,q} \right) \right| - (p+q) + O_p(1), \quad (5.6)$$

2. si los parámetros son evaluados en $\hat{\beta}_{p,q}$ y $\hat{\sigma}_{p,q}^2$:

$$-\frac{T}{2}(\log 2\pi + 1) - \frac{T}{2} \log \hat{\sigma}_{p,q}^2 - \frac{1}{2} \log \left| Q_T \left(\hat{\beta}_{p,q} \right) \right| - \frac{T(p+q+1)}{(T-p-q-2)} + O_p(1). \quad (5.7)$$

La aproximación (5.6) lleva al criterio,

$$AIC^*(p, q) = \log \hat{\sigma}_{p,q}^2 + \frac{2(p+q)}{T} + \frac{\log \left| Q_T \left(\hat{\beta}_{p,q} \right) \right|}{T} = AIC(p, q) + \frac{\log \left| Q_T \left(\hat{\beta}_{p,q} \right) \right|}{T} \quad (5.8)$$

que es la versión corregida del criterio de Akaike. La expresión (5.6) lleva al criterio de selección de modelos de la forma,

$$AICc^*(p, q) = \log \hat{\sigma}_{p,q}^2 + \frac{2(p+q+1)}{(T-p-q-2)} + \frac{\log \left| Q_T \left(\hat{\beta}_{p,q} \right) \right|}{T} = AICc(p, q) + \frac{\log \left| Q_T \left(\hat{\beta}_{p,q} \right) \right|}{T} \quad (5.9)$$

que es la versión corregida del criterio de Hurvich y Tsai.

5.2.2. Una aproximación Bayesiana

A continuación, se analiza la regla en (5.3) teniendo en cuenta que este enfoque requiere probabilidades a priori de los modelos, $p(M_{p,q})$ y de los parámetros, $p(\alpha_{p,q}|M_{p,q})$. El punto de vista Bayesiano de maximizar la probabilidad a posteriori ha sido ampliamente considerado, ver Schwarz (1978), Chow (1981), Haughton (1988) o Raftery, Madigan and Volinsky (1995), pero, que el autor de esta Tesis sepa, no ha sido formalmente analizado como un problema de discriminación. La demostración del siguiente Lema se presenta en el apéndice.

Lema 5.2 *Bajo la hipótesis 1, el logaritmo de la probabilidad de x dados los parámetros está dado por,*

$$\begin{aligned} \log p(x|M_{p,q}) &= \frac{1}{2}(p+q+1-T) \log(2\pi) - \frac{1}{2}(p+q+1) \log T - \frac{T}{2} \log \hat{\sigma}_{p,q}^2 \\ &\quad - \frac{1}{2} \log \left| Q_T \left(\hat{\beta}_{p,q} \right) \right| - \frac{1}{2}T + \log p(\hat{\alpha}_{p,q}|M_{p,q}) + O_p(1). \end{aligned} \quad (5.10)$$

Entonces, dividiendo por T en (5.10), tomando las mismas probabilidades a priori para todos los parámetros e ignorando algunos términos constantes, se obtiene el criterio de selección BIC corregido,

$$BIC^*(p, q) = \log \hat{\sigma}_{p,q}^2 + \frac{\log(T)(p+q+1)}{T} + \frac{\log \left| Q_T \left(\hat{\beta}_{p,q} \right) \right|}{T} = BIC(p, q) + \frac{\log \left| Q_T \left(\hat{\beta}_{p,q} \right) \right|}{T}. \quad (5.11)$$

Todos los criterios obtenidos pueden ser escritos de manera compacta como miembros de la familia de criterios,

$$\min_{(p,q)} \left\{ \log \hat{\sigma}_{p,q}^2 + (p+q) \times C(T, p+q+1) + \frac{\log |Q_T(\hat{\beta}_{p,q})|}{T} \right\}, \quad (5.12)$$

donde el término $|Q_T(\hat{\beta}_{p,q})|$ se calcula como se detallará en la sección 5.3. Las propiedades de eficiencia y consistencia de los criterios en (5.12) no están afectadas por el término de corrección, como ilustra el siguiente Teorema cuya prueba aparece en el apéndice de este capítulo, y que demuestra que los criterios (5.8) y (5.9) son eficientes.

Teorema 5.3 *Bajo las hipótesis: (A1) $\{x_t\}$ es generada por un proceso estacionario $x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots = a_t$, $t = \dots, -1, 0, 1, \dots$ donde a_t es una secuencia de variables aleatorias Gaussianas independientes con media 0 y varianza σ_a^2 y $\sum_{j=1}^{\infty} |\phi_j| < \infty$; (A2) La secuencia $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots$, es no cero para $|B| \leq 1$; (A3) p_{\max} es una secuencia de enteros positivos tales que $p_{\max} \rightarrow \infty$ y $p_{\max}/\sqrt{T} \rightarrow 0$ cuando $T \rightarrow \infty$; (A4) $\{x_t\}$ no es degenerada a un proceso autorregresivo de orden finito, los criterios AIC^* y $AICc^*$ son eficientes.*

De este modo, AIC^* y $AICc^*$ son similares a AIC y $AICc$ para muestras grandes, pero incluyen el factor de corrección $\log |Q_T(\hat{\beta}_{p,q})|/T$, que como se verá es importante en muestras finitas. Para el criterio en (5.11), la propiedad de consistencia se mantiene gracias al Teorema 3 en Hannan (1980, p. 1073). Entonces, el criterio (5.11) es un criterio consistente.

Hay que remarcar que la principal contribución de esta sección es ver el problema de selección de modelos como un análisis de discriminación y presentar un enfoque unificado de los criterios propuestos en la literatura desde diferentes puntos de vista. La conexión entre estos dos problemas de decisión no ha sido previamente planteada. Los detalles técnicos en los enfoques de máxima verosimilitud y Bayesiano se incluyen para completar los razonamientos. La inclusión del término de corrección fue analizada por Hannan (1980) quien concluyó que puede ser omitido porque tiende a cero. Aquí se demuestra en la sección 5.5 su importancia en series de datos finitos. Hannan (1980) omite este término al derivar las propiedades de consistencia del BIC. El Teorema 5.3 demuestra que los resultados en Hannan (1980) y Shibata (1980) no están afectados por la inclusión de este término.

5.3. Análisis del término de corrección

Un punto clave en la discusión de la sección anterior es obtener una expresión manejable del término $|Q_T(\hat{\beta}_{p,q})|$. Este determinante depende de los estimadores máximo verosímiles

bajo la hipótesis de un modelo ARMA(p, q). Leeuw (1994) obtiene una expresión para $|Q_T(\beta_{p,q})|$ que sólo depende de los parámetros $\beta_{p,q}$ y sólo requiere el cálculo de dos determinantes de orden $m = \max(p, q)$. Dicha expresión puede ser escrita como,

$$|Q_T(\beta_{p,q})| = \frac{|(R'R - SS') + (RV - US)' H_1' H_1 (RV - US)|}{|R'R - SS'|}, \quad (5.13)$$

donde R, S, U y V son matrices de tamaño $m \times m$ dadas por,

$$R = \begin{cases} 0 & i < j \\ 1 & i = j \\ -\phi_{i-j} & i > j \end{cases} \quad U = \begin{cases} 0 & i < j \\ 1 & i = j \\ -\theta_{i-j} & i > j \end{cases}$$

$$S = \begin{cases} -\phi_{m+(i-j)} & i \leq j \\ 0 & i > j \end{cases} \quad V = \begin{cases} -\theta_{m+(i-j)} & i \leq j \\ 0 & i > j \end{cases}$$

y la matriz H_1 es la matriz de tamaño $T \times m$ definida por las primeras m columnas de la matriz adjunta H de tamaño $T \times T$, que tiene la misma estructura que U con ceros en cualquier posición que no sean las primeras m diagonales inferiores. Los elementos ϕ_i de las matrices previas son 0 si $m > p$ y los elementos θ_i son 0 si $m > q$. Por lo tanto, se puede obtener $|Q_T(\hat{\beta}_{p,q})|$ substituyendo los estimadores $\hat{\beta}_{p,q}$ en las fórmulas de las matrices R, S, U, V y H .

A continuación, analizamos el término $|Q_T(\beta_{p,q})|$. Como $Q_T(\beta_{p,q}) = (\sigma_x^2/\sigma_a^2) R_T(\alpha_{p,q})$, donde σ_x^2 es la varianza de x_t y $R_T(\alpha_{p,q})$ es la matriz de correlación de orden T , se tiene que $|Q_T(\beta_{p,q})| = (\sigma_x^2/\sigma_a^2)^T |R_T(\alpha_{p,q})|$. Durbin (1960) y Ramsey (1974) demuestran que,

$$\sigma_x^2/\sigma_a^2 = \prod_{i=1}^{\infty} (1 - \phi_{ii}^2(\beta_{p,q}))^{-1}, \quad |R_T(\alpha_{p,q})| = \prod_{i=1}^{T-1} (1 - \phi_{ii}^2(\beta_{p,q}))^{T-i}$$

respectivamente, donde $\phi_{ii}(\beta_{p,q})$ son las correlaciones parciales del proceso. Es importante notar que estas correlaciones parciales se obtienen bajo la hipótesis de un modelo ARMA(p, q). Entonces,

$$|Q_T(\beta_{p,q})| = \frac{\prod_{i=1}^{T-1} (1 - \phi_{ii}^2(\beta_{p,q}))^{T-i}}{\prod_{i=1}^{\infty} (1 - \phi_{ii}^2(\beta_{p,q}))^T} = \prod_{i=1}^{T-1} (1 - \phi_{ii}^2(\beta_{p,q}))^{-i} \prod_{i=T}^{\infty} (1 - \phi_{ii}^2(\beta_{p,q}))^{-T} \quad (5.14)$$

y los criterios (5.12) se pueden escribir como sigue:

$$\min_{(p,q)} \left\{ \log \hat{\sigma}_{p,q}^2 + (p+q) \times C(T, p+q) - \sum_{i=1}^{T-1} \frac{i}{T} \log \left(1 - \phi_{ii}^2(\hat{\beta}_{p,q}) \right) - \sum_{i=T}^{\infty} \log \left(1 - \phi_{ii}^2(\hat{\beta}_{p,q}) \right) \right\}.$$

Para entender mejor los criterios corregidos, la diferencia $AIC^*(p+1, q) - AIC^*(p, q)$ está dada por,

$$\begin{aligned} AIC^*(p+1, q) - AIC^*(p, q) = & \log \frac{\hat{\sigma}_{p+1, q}^2}{\hat{\sigma}_{p, q}^2} + \frac{2}{T} - \sum_{i=1}^{T-1} \frac{i}{T} \log \left(1 - \phi_{ii}^2(\hat{\beta}_{p+1, q}) \right) - \\ & - \sum_{i=T}^{\infty} \log \left(1 - \phi_{ii}^2(\hat{\beta}_{p+1, q}) \right) + \sum_{i=1}^{T-1} \frac{i}{T} \log \left(1 - \phi_{ii}^2(\hat{\beta}_{p, q}) \right) + \sum_{i=T}^{\infty} \log \left(1 - \phi_{ii}^2(\hat{\beta}_{p, q}) \right), \end{aligned}$$

y que puede ser aproximada por,

$$\frac{\hat{\sigma}_{p+1, q}^2 - \hat{\sigma}_{p, q}^2}{\hat{\sigma}_{p, q}^2} + \frac{2}{T} + \sum_{i=1}^{T-1} \frac{i}{T} \left(\phi_{ii}^2(\hat{\beta}_{p+1, q}) - \phi_{ii}^2(\hat{\beta}_{p, q}) \right) + \sum_{i=T}^{\infty} \left(\phi_{ii}^2(\hat{\beta}_{p+1, q}) - \phi_{ii}^2(\hat{\beta}_{p, q}) \right).$$

El primer término mide el cambio relativo entre las varianzas, $\hat{\sigma}_{p+1, q}^2$ y $\hat{\sigma}_{p, q}^2$. El segundo término es una penalización por la inclusión de un parámetro adicional. Los términos tercero y cuarto miden la discrepancia entre todos los coeficientes de autocorrelación bajo ambas hipótesis, $ARMA(p, q)$ y $ARMA(p+1, q)$, con pesos que se incrementan con el retardo. Entonces, $AIC^*(p+1, q) < AIC^*(p, q)$ si: (a) $\hat{\sigma}_{p+1, q}^2$ es significativamente menor que $\hat{\sigma}_{p, q}^2$, o (b) la suma ponderada de coeficientes de correlación parcial obtenidos bajo el modelo $ARMA(p, q)$ es mayor que la correspondiente suma bajo el modelo $ARMA(p+1, q)$. La misma interpretación se aplica al BIC^* y $AICc^*$, con la única diferencia del término de penalización por incluir un parámetro adicional.

En el caso de ajuste autorregresivo, donde los criterios (5.12) tienen la forma:

$$\min_p \left\{ \log \hat{\sigma}_p^2 + p \times C(T, p) + \frac{\log |Q_T(\hat{\beta}_p)|}{T} \right\}, \quad (5.15)$$

el término (5.13) se reduce a:

$$|Q_T(\beta_p)| = \frac{1}{|R'R - SS'|} \quad (5.16)$$

y utilizando que $\sigma_x^2/\sigma_a^2 = \prod_{i=1}^p (1 - \phi_{ii}^2(\beta_p))^{-1}$ y teniendo en cuenta que $\phi_{ii}^2(\beta_p) = 0$ si $i > p$,

$$|Q_T(\beta_p)| = \frac{\prod_{i=1}^{T-1} (1 - \phi_{ii}^2(\beta_p))^{T-i}}{\prod_{i=1}^p (1 - \phi_{ii}^2(\beta_p))^T} = \prod_{i=1}^p (1 - \phi_{ii}^2(\beta_p))^{-i}. \quad (5.17)$$

Por lo tanto, para modelos $AR(p)$, los criterios (5.15) se pueden escribir como sigue:

$$\min_p \left\{ \log \hat{\sigma}_p^2 + p \times C(T, p) - \sum_{i=1}^p \frac{i}{T} \log \left(1 - \phi_{ii}^2(\hat{\beta}_p) \right) \right\},$$

y la diferencia $AIC^*(p+1) - AIC^*(p)$ puede ser aproximada por:

$$\frac{\hat{\sigma}_{p+1}^2 - \hat{\sigma}_p^2}{\hat{\sigma}_p^2} + \frac{2}{T} + \sum_{i=1}^p \frac{i}{T} \left(\phi_{ii}^2(\hat{\beta}_{p+1}) - \phi_{ii}^2(\hat{\beta}_p) \right) + \frac{p+1}{T} \phi_{p+1,p+1}^2(\hat{\beta}_{p+1}).$$

Como en el caso de modelos ARMA, el primer término mide el cambio relativo entre las varianzas, $\hat{\sigma}_p^2$ y $\hat{\sigma}_{p+1}^2$, el segundo término es la penalización por incluir un parámetro adicional y el tercero mide la discrepancia entre las primeras p autocorrelaciones parciales bajo ambas hipótesis, $AR(p)$ y $AR(p+1)$. Finalmente, el último término mide la significatividad del coeficiente de autocorrelación de orden $p+1$.

5.4. Criterios de selección en modelos TAR

Uno de los modelos más utilizados para series temporales no lineales es el modelo autorregresivo por umbrales (TAR). Una serie temporal, $x = (x_1, \dots, x_T)'$ ha sido generada por la clase de procesos TAR si sigue el modelo:

$$x_t = \phi_{j0} + \sum_{i=1}^{p_j} \phi_{ji} x_{t-i} + a_{jt}, \quad \text{if } r_{j-1} \leq x_{t-d} < r_j \quad (5.18)$$

donde se supone que a_{jt} , $j = 1, \dots, k$, es una serie de ruido blanco con media 0 y varianzas finitas $\sigma_{a_j}^2$, $d \in \{0, \dots, d_{\max}\}$ es un entero positivo y $-\infty = r_0 < r_1 < \dots < r_{k-1} < r_k = \infty$ son los umbrales. También se supone que $p_j \in \{0, \dots, p_j^{\max}\}$ donde p_j^{\max} con cotas máximas fijas. Se denota por $M_{p_1, \dots, p_k, d}$ al modelo $TAR(p_1, \dots, p_k, d)$.

Como se menciona en el capítulo 2, no se consideran estimadores máximo verosímiles exactos de los parámetros del modelo (5.18) porque los umbrales r_1, \dots, r_{k-1} no son parámetros continuos. Entonces, los parámetros del modelo (5.18) son estimados por verosimilitud condicional. Se supone la siguiente hipótesis:

Hipótesis 2: Los modelos $M_{p_1, \dots, p_k, d}$ son estacionarios y ergódicos con momentos segundos finitos y la distribución estacionaria de $x = (x_1, \dots, x_T)'$ admite una densidad que es positiva en todo punto.

Chan (1993) demuestra que bajo la hipótesis 2, los estimadores mínimo cuadráticos condicionales de los parámetros de un modelo autorregresivo por umbrales son fuertemente consistentes. Sea $\phi_j = (\phi_{j0}, \dots, \phi_{jp_j})'$, $j = 1, \dots, k$, el vector de parámetros. La función suma de cuadrados se define por:

$$\begin{aligned} S_x(\phi_1, \dots, \phi_k, r_0, \dots, r_k) &= S_x(\phi_1, r_0, r_1) + \dots + S_x(\phi_k, r_{k-1}, r_k) = \\ &= \sum_{r_0 < x_{t-d} \leq r_1} a_t^2 + \dots + \sum_{r_{k-1} \leq x_{t-d} < r_k} a_t^2, \end{aligned} \quad (5.19)$$

y los estimadores mínimo cuadráticos condicionales de $(\phi_1, \dots, \phi_k, r_1, \dots, r_{k-1})$ son los valores que minimizan (5.19), que se denotan por $\hat{\phi}_1, \dots, \hat{\phi}_k$ y $\hat{r}_1, \dots, \hat{r}_{k-1}$. Las varianzas residuales se definen como,

$$\hat{\sigma}_1^2 = \frac{S_x(\hat{\phi}_1, r_0, \hat{r}_1)}{T_1}, \quad \hat{\sigma}_j^2 = \frac{S_x(\hat{\phi}_j, \hat{r}_{j-1}, \hat{r}_j)}{T_j}, \quad j = 2, \dots, k-1, \quad \hat{\sigma}_k^2 = \frac{S_x(\hat{\phi}_k, \hat{r}_{k-1}, r_k)}{T_k}$$

donde T_j son el número de observaciones en cada regimen para los estimadores $\hat{r}_1, \dots, \hat{r}_{k-1}$.

Se ha prestado poca atención a los criterios de selección de modelos TAR. Wong y Li (1998) derivan el criterio AICc para el caso en que $k = 2$, y proponen un procedimiento para seleccionar y estimar un modelo TAR. También comparan mediante un estudio de simulación tres criterios de selección de modelos, AIC, AICc y BIC, que para k regimenes están dados por:

$$\begin{aligned} BIC(p_1, \dots, p_k) &= \sum_{j=1}^k \{T_j \log \hat{\sigma}_j^2 + \log T_j (p_j + 1)\} \\ AIC(p_1, \dots, p_k) &= \sum_{j=1}^k \{T_j \log \hat{\sigma}_j^2 + 2(p_j + 2)\} \\ AICc(p_1, \dots, p_k) &= \sum_{j=1}^k \left\{ T_j \log \hat{\sigma}_j^2 + \frac{T_j (T_j + p_j + 1)}{T_j - p_j - 3} \right\}. \end{aligned} \quad (5.20)$$

El procedimiento propuesto por Wong y Li (1998) cuando $k = 2$ y $r_1 = r$ y d son desconocidos es el siguiente: (a) Fijar los ordenes autorregresivos y del retardo máximos $\{p_1^{\max}, p_2^{\max}, d^{\max}\}$; (b) Suponer $r \in [l, u] \subset R$, donde l es el percentil $0.25 \times 100\%$ y u es el percentil $0.75 \times 100\%$ de x_t ; (c) Sea $x_{(1)}, \dots, x_{(T)}$ los estadísticos del orden de x_t ; (d) Sea $I_r = \{[0.25T], \dots, [0.75T]\}$. Sea $r = x_{(i)}$, $i \in I_r$; (e) Calcular,

$$\min \{C(p_1, p_2, d, x_{(i)}) : p_1 \in \{1, \dots, p_1^{\max}\}, p_2 \in \{1, \dots, p_2^{\max}\}, d \in \{1, \dots, d^{\max}\}, x_{(i)} \in I_r\},$$

donde $C(p_1, p_2, d, x_{(i)})$ es el criterio de selección utilizado. Los ordenes autorregresivos y el parámetro de retardo, $(p_1, p_2, d, x_{(i)})$ y d , respectivamente y el umbral estimado son los que minimizan $C(p_1, p_2, d, x_{(i)})$.

Este procedimineto proporciona los ordenes autorregresivos y el parámetro retardo seleccionado por el criterio y una estimación del umbral. Wong y Li (1998) llevan a cabo un experimento de Monte Carlo para diferentes modelos y tamaños muestrales para los criterios en (5.20), y concluye que el criterio AICc es adecuado para tamaños muestrales pequeños pero el criterio BIC es preferible para tamaños muestrales medios y largos.

De Gooijer (2001) propone un procedimiento para seleccionar y estimar los parámetros de un modelo TAR para tres criterios de validación cruzada. Los primeros cuatro pasos del procedimiento propuesto para el primer criterio que se denota por C_1 son como en Wong y

Li (1998). El resto del procedimiento es como sigue: (e) Omitir una observación de la serie y con las restantes obtener el estimador mínimo cuadrático condicional de los parámetros del modelo correspondiente, que se denota por $\hat{\phi}_j^t$, predecir la observación omitida y obtener el residuo predictivo, $a_t(\hat{\phi}_j^t, \hat{r}_{j-1}, \hat{r}_j)$; (f) Repetir el paso anterior para todas las observaciones y el modelo final es aquel que minimiza el criterio C_1 definido como sigue:

$$C_1(p_1, \dots, p_k) = T \log \left(T^{-1} \sum_{t=1}^T a_t^2(\hat{\phi}_j^t, \hat{r}_{j-1}, \hat{r}_j) \right). \quad (5.21)$$

Stoica et al (1986) demuestran que para un modelo dado,

$$C_1(p_1, \dots, p_k) = AIC(p_1, \dots, p_k) + O(T^{-\frac{1}{2}}). \quad (5.22)$$

y basado en la definición del criterio AICc y (5.22), De Gooijer (2001) define el criterio siguiente:

$$Cc(p_1, \dots, p_k) = C_1(p_1, \dots, p_k) + \sum_{j=1}^k \left\{ \frac{T_j(T_j + p_j + 1)}{T_j - p_j - 3} \right\}. \quad (5.23)$$

De Gooijer (2001) propone también una generalización de un criterio de selección de modelos introducido por McQuarrie, Shumway y Tsai (1997) para modelos lineales. Este criterio no es eficiente ni consistente, pero tiene buenas propiedades para muestras pequeñas, y está dado por,

$$AICu(p_1, \dots, p_k) = AICc(p_1, \dots, p_k) + \sum_{j=1}^k T_j \log \left\{ \frac{T_j}{T_j - p_j - 2} \right\}, \quad (5.24)$$

mientras que el criterio de validación cruzada propuesto por De Gooijer (2001) es de la forma:

$$Cu(p_1, \dots, p_k) = Cc(p_1, \dots, p_k) + \sum_{j=1}^k T_j \log \left\{ \frac{T_j}{T_j - p_j - 2} \right\}. \quad (5.25)$$

Como los modelos TAR son lineales por tramos, se puede incluir el término de corrección en (5.16) en cada uno de los regímenes. Por lo tanto, se explorará el comportamiento del procedimiento de Wong y Li (1998) pero modificando los criterios BIC, AIC, AICc y AICu por:

$$C^*(p_1, \dots, p_k) = C(p_1, \dots, p_k) + \sum_{j=1}^k \log |Q_{T_j}(\hat{\phi}_j)|, \quad (5.26)$$

donde $C(p_1, \dots, p_k)$ representa los criterios BIC, AIC, AICc y AICu en (5.20) y (5.24) respectivamente. Para obtener el término de corrección en cada régimen, primero se estiman



los parámetros del modelo por máxima verosimilitud condicional y se obtiene el término de corrección en cada regimen como en (5.16).

De la misma manera, los criterios de validación cruzada propuesto por De Gooijer (2001) se pueden modificar como sigue:

$$C_1^*(p_1, \dots, p_k) = C_1(p_1, \dots, p_k) + \sum_{j=1}^k \log \left| Q_{T_j}(\hat{\phi}_j) \right|, \quad (5.27)$$

donde $C(p_1, \dots, p_k)$ representa $C_1(p_1, \dots, p_k)$, $C_c(p_1, \dots, p_k)$ y $C_u(p_1, \dots, p_k)$. En este caso, el procedimiento en De Gooijer (2001) se modifica añadiendo el término de corrección en el último paso estimando los parámetros por máxima verosimilitud condicional con todas las observaciones de la serie. Entonces, el modelo final seleccionado es aquel que minimiza uno de los criterios (5.27). El término de corrección se obtiene como en el caso previo.

5.5. Experimentos de Monte Carlo

5.5.1. Experimentos para modelos ARMA

Para evaluar el comportamiento de los criterios propuestos para modelos ARMA y diferentes tamaños muestrales, se generan 1000 series para cada uno de los siguientes seis modelos, (M1) $x_t = 0.9x_{t-1} + a_t$, (M2) $x_t = 1.4x_{t-1} - 0.7x_{t-2} + a_t$, (M3) $x_t = 0.4x_{t-1} - 0.8x_{t-2} + 0.6x_{t-3} + a_t$, (M4) $x_t = a_t + 0.8a_{t-1}$, (M5) $x_t = 0.8x_{t-1} + a_t + 0.7a_{t-1}$ and (M6) $x_t = 1.4x_{t-1} - 0.7x_{t-2} + a_t + 0.8a_{t-1}$, donde a_t son normales estándar independientes. Estos modelos han sido elegidos para representar diferentes situaciones. Los primeros tres representan estructuras autorregresivas comunes: una fuerte dependencia AR(1) (M1), un AR(2) con dos raíces complejas (M2), y un modelo AR(3) con una raíz real y dos complejas (M3). Los segundos tres modelos incluyen modelos ARMA que requieren aproximaciones autorregresivas largas con raíces reales (M4 y M5) y mixturas de raíces real y compleja (M6). Los tres primeros modelos se utilizan para demostrar el comportamiento de los criterios corregidos en modelos de ordenes autorregresivos finitos y los últimos tres en procesos ARMA. Basados en las secciones previas, se compara el comportamiento de criterios de la forma:

$$\min_{p,q} \{ \log \hat{\sigma}_{p,q}^2 + (p+q) \times C(T, p+q) \}$$

y,

$$\min_{p,q} \left\{ \log \hat{\sigma}_{p,q}^2 + (p+q) \times C(T, p+q) + \frac{\log \left| Q_T(\hat{\beta}_{p,q}) \right|}{T} \right\}.$$

En todos los casos, se generan 1000 series de cada modelo, M1 a M6 con tamaños muestrales $T = 31, 51$ y 101 . Para procesos autorregresivos, donde $q = 0$, se ajusta cada modelo

Tabla 5.1: Frecuencia de selección del modelo correcto y raíz del error cuadrático medio de predicción para los modelos seleccionados por cada criterio. Si C es un criterio dado, C* es la versión corregida incluyendo el término de corrección propuesto en esta Tesis.

<i>T</i> = 30													
M	BIC	BIC*	AIC	AIC*	AICc	AICc*	M	BIC	BIC*	AIC	AIC*	AICc	AICc*
1	793	851	469	612	667	749	1	1.28	1.16	1.58	1.35	1.17	1.12
2	763	829	454	662	694	782	2	1.13	1.08	1.41	1.26	1.03	0.99
3	566	587	422	578	596	623	3	1.22	1.10	1.39	1.23	1.07	1.03
4	495	595	187	309	360	460	4	1.03	1.01	1.12	1.05	1.07	1.02
5	302	423	101	185	224	352	5	1.09	1.05	1.10	1.02	1.09	1.01
6	314	510	111	270	262	474	6	1.04	0.99	1.07	1.03	1.04	1.03
<i>T</i> = 50													
M	BIC	BIC*	AIC	AIC*	AICc	AICc*	M	BIC	BIC*	AIC	AIC*	AICc	AICc*
1	879	902	538	641	675	727	1	1.06	1.06	1.10	1.10	1.09	1.09
2	878	914	553	677	695	771	2	1.06	1.06	1.10	1.08	1.08	1.07
3	802	819	554	681	690	745	3	1.11	1.11	1.15	1.13	1.12	1.11
4	655	728	246	314	326	399	4	1.09	1.09	1.24	1.16	1.17	1.14
5	554	628	226	283	314	397	5	1.13	1.13	1.18	1.16	1.17	1.14
6	598	725	265	384	385	512	6	1.03	1.00	1.06	1.01	1.01	1.00
<i>T</i> = 100													
M	BIC	BIC*	AIC	AIC*	AICc	AICc*	M	BIC	BIC*	AIC	AIC*	AICc	AICc*
1	931	940	584	629	633	671	1	1.17	1.17	1.17	1.16	1.16	1.16
2	906	917	589	655	663	704	2	0.99	0.98	0.94	0.94	0.93	0.90
3	944	954	622	702	685	758	3	1.09	1.09	1.12	1.10	1.09	1.10
4	833	852	378	418	441	458	4	1.14	1.14	1.23	1.23	1.23	1.22
5	826	843	418	435	463	493	5	0.91	0.89	0.92	0.92	0.92	0.92
6	857	888	500	537	564	607	6	1.02	1.01	1.03	1.03	1.04	1.03

con las primeras $T - 1$ observaciones de cada serie por máxima verosimilitud y en cada modelo, se obtiene $\left| Q_T \left(\hat{\beta}_p \right) \right|$ como en (5.16). Se fija $p_{\max} = 15$, por lo que se ajustan 16 modelos para cada serie. Para procesos ARMA se ajustan los modelos del conjunto de $(p_{\max} + 1) \times (q_{\max} + 1)$ ordenes ARMA para las primeras $T - 1$ observaciones donde $p_{\max} = 4$ y $q_{\max} = 4$ por estimación máximo verosímil, por lo que se ajustan 25 modelos para cada serie. En cada modelo, se obtiene $\left| Q_T \left(\hat{\beta}_{p,q} \right) \right|$ mediante (5.13). En ambos casos, con el modelo seleccionado por cada criterio y los parámetros ajustados, se obtiene la predicción para la última observación y se compara con el verdadero valor de la observación. Como los criterios BIC y BIC* son consistentes, esto es, buscan determinar el verdadero modelo, y AIC, AICc, AIC* y AICc* son criterios eficientes, esto es, buscan el modelo que mejor predice, se analizan las propiedades de consistencia y eficiencia para todos los criterios en muestras pequeñas y medianas.

Es importante hacer notar que el ajuste de los $(p_{\max} + 1) \times (q_{\max} + 1)$ modelos ARMA puede llevar a serios problemas de estimación. Ver Hannan y Rissanen (1982), Hannan y Kavalieris (1984), Poskitt (1987), Pukkila, Koreisha y Kallinen (1990) o Pötscher (1990), entre otros. Estos autores han propuesto algoritmos para estimar los ordenes (p, q) de un proceso ARMA mediante la estimación consistente de $m = \max(p, q)$. Conocer m elimina la estimación de procesos ARMA sobreparametrizados lo que lleva a problemas de inconsistencia. Aquí, para simplificar, se considera la búsqueda total sobre todos los modelos candidatos. Cuando se encuentra una matriz de covarianzas singular en el experimento de Monte Carlo, se elimina de la comparación con otros modelos.

Se comparan los tres criterios con sus versiones corregidas en los seis modelos, por lo que se llevan a cabo 54 comparaciones. Las frecuencias sobre 1000 que los criterios seleccionan el orden correcto para los seis modelos se muestran en las columnas 2 a 7 en la Tabla 5.1. Se puede ver que para tamaños muestrales pequeños, $T = 30$, la mejora en el número de veces que se selecciona el verdadero modelo puede ser tan grande como un 143 % (ver AIC y AIC* en M6) y para $T = 100$ puede ser del 13 % (ver AIC y AIC* en M3). Por otro lado, las columnas 9 a 14 en la Tabla 5.1 muestran la raíz del error cuadrático medio de predicción estimado para cada criterio en todos los tamaños muestrales. Para $T = 30$, los criterios corregidos mejoran el comportamiento predictivo de los criterios originales. Para $T = 50$ y $T = 100$, los errores cuadráticos medios de predicción de los criterios corregidos son también menores en la mayoría de los casos, pero las diferencias entre los criterios originales y corregidos es pequeña, especialmente para $T = 100$.

La Tabla 5.2 presenta la media y el error estándar de los ordenes seleccionados por cada criterio para ajuste autorregresivo para los primeros tres modelos. Para todos los tamaños muestrales considerados, los criterios corregidos se comportan mejor que los originales. Notar que la inclusión del término $\left| Q_T \left(\hat{\beta}_p \right) \right|$ es muy efectiva en reducir el error estándar del orden seleccionado por los criterios, especialmente para el caso de $T = 30$, donde la desviación se reduce en más del 50 % por la inclusión del término de corrección.

Tabla 5.2: Media y error estándar de los ordenes seleccionados por los criterios para M1, M2 y M3.

$T = 30$						
M	BIC	BIC*	AIC	AIC*	AICc	AICc*
1	1.50 (1.54)	1.23 (0.73)	3.93 (4.26)	2.04 (1.94)	1.71 (1.41)	1.41 (0.94)
2	2.57 (1.71)	2.19 (0.68)	4.86 (3.99)	2.83 (1.69)	2.60 (1.28)	2.31 (0.81)
3	3.23 (1.69)	2.80 (0.74)	5.12 (3.62)	3.30 (1.42)	3.19 (1.15)	2.89 (0.76)
$T = 50$						
M	BIC	BIC*	AIC	AIC*	AICc	AICc*
1	1.20 (0.70)	1.13 (0.48)	3.27 (3.69)	2.13 (2.23)	1.87 (1.81)	1.56 (1.24)
2	2.18 (0.66)	2.10 (0.43)	4.01 (3.39)	2.90 (1.87)	2.78 (1.67)	2.44 (1.05)
3	3.15 (0.77)	3.04 (0.56)	4.53 (2.74)	3.64 (1.52)	3.56 (1.32)	3.31 (0.92)
$T = 100$						
M	BIC	BIC*	AIC	AIC*	AICc	AICc*
1	1.09 (0.40)	1.07 (0.33)	2.66 (3.00)	2.20 (2.33)	2.13 (2.15)	1.86 (1.72)
2	2.12 (0.44)	2.10 (0.36)	3.69 (2.90)	3.13 (2.18)	3.04 (2.01)	2.76 (1.63)
3	3.07 (0.36)	3.05 (0.31)	4.38 (2.59)	3.86 (1.89)	3.89 (1.87)	3.56 (1.32)

Tabla 5.3: Raíz del error cuadrático medio de predicción para M4.

T	BIC	BIC*	AIC	AIC*	AICc	AICc*
30	1.33	1.23	1.51	1.36	1.13	1.11
50	1.15	1.14	1.17	1.14	1.12	1.13
100	1.09	1.10	1.09	1.08	1.09	1.09

Tabla 5.4: Frecuencia de selección del orden correcto, raíz del error cuadrático medio de estimación y raíz del error cuadrático medio de predicción suponiendo que el retardo es conocido.

M	$T = 30$	BIC	BIC*	AIC	AIC*	AICc	AICc*	AICu	AICu*	C_1	C_1^*	Cc	Cc*	Cu	Cu*
7	(p_1, p_2)	306	377	254	331	800	856	903	923	363	474	818	854	895	923
7	RMSE	0.62	0.62	0.60	0.60	0.41	0.40	0.40	0.39	0.55	0.54	0.40	0.40	0.40	0.39
7	RMSPE	1.41	1.37	1.38	1.32	1.13	1.13	1.12	1.12	1.22	1.22	1.14	1.12	1.12	1.11
8	(p_1, p_2)	306	361	243	310	779	825	876	913	378	450	786	840	890	921
8	RMSE	0.84	0.84	0.82	0.82	0.65	0.66	0.65	0.65	0.73	0.74	0.64	0.65	0.63	0.63
8	RMSPE	1.64	1.63	1.64	1.64	1.16	1.15	1.13	1.12	1.31	1.31	1.14	1.13	1.12	1.12
M	$T = 50$	BIC	BIC*	AIC	AIC*	AICc	AICc*	AICu	AICu*	C_1	C_1^*	Cc	Cc*	Cu	Cu*
7	(p_1, p_2)	420	512	198	298	629	686	789	827	286	395	634	688	796	836
7	RMSE	0.58	0.56	0.55	0.54	0.47	0.47	0.47	0.47	0.53	0.52	0.45	0.45	0.45	0.44
7	RMSPE	1.15	1.14	1.19	1.19	1.06	1.06	1.06	1.06	1.15	1.15	1.07	1.07	1.07	1.06
8	(p_1, p_2)	412	507	216	307	633	704	802	846	313	406	659	724	820	855
8	RMSE	0.74	0.74	0.73	0.74	0.63	0.63	0.63	0.64	0.70	0.70	0.63	0.65	0.64	0.64
8	RMSPE	1.19	1.19	1.24	1.23	1.15	1.15	1.13	1.12	1.22	1.22	1.15	1.14	1.12	1.12
M	$T = 100$	BIC	BIC*	AIC	AIC*	AICc	AICc*	AICu	AICu*	C_1	C_1^*	Cc	Cc*	Cu	Cu*
7	(p_1, p_2)	743	796	359	431	537	591	753	789	358	439	522	589	765	799
7	RMSE	0.54	0.52	0.52	0.51	0.50	0.48	0.50	0.47	0.51	0.49	0.49	0.47	0.48	0.46
7	MSPE	1.06	1.06	1.08	1.06	1.07	1.06	1.06	1.06	1.08	1.08	1.07	1.07	1.06	1.06
8	(p_1, p_2)	771	819	365	446	545	600	773	816	398	461	542	611	773	805
8	RMSE	0.59	0.59	0.58	0.59	0.57	0.58	0.57	0.57	0.58	0.58	0.56	0.57	0.56	0.57
8	RMSPE	1.06	1.05	1.07	1.07	1.07	1.07	1.06	1.05	1.07	1.07	1.07	1.06	1.06	1.06

Por último, se realiza un último experimento para analizar el comportamiento predictivo de aproximaciones AR a modelos ARMA. Se generan 1000 series del modelo M4 para tamaños muestrales $T = 31, 51$ y 101 . Se ajusta un modelo autorregresivo por máxima verosimilitud con $p_{\max} = 15$ para las primeras $T - 1$ observaciones, y se obtiene el error de predicción para la última observación por el modelo seleccionado por cada criterio. Con las 1000 series se estima la raíz del error cuadrático medio de predicción. La Tabla 3 muestra los resultados. Para $T = 30$, en todos los casos, los criterios corregidos mejoran los criterios originales. Para los tamaños muestrales $T = 50$ y $T = 100$, los criterios corregidos mejoran el comportamiento de los criterios originales en los casos de BIC y AIC, pero las diferencias entre ellos es muy pequeña. Se concluye que el término de corrección mejora el comportamiento en muestras pequeñas y medianas para los criterios corregidos.

5.5.2. Experimentos para modelos TAR

Para evaluar el comportamiento de los criterios propuestos para modelos TAR en diferentes modelos y tamaños muestrales, se generan 1000 series de los siguientes dos modelos

Tabla 5.5: Frecuencia de selecci3n del orden correcto, ra3z del error cuadr3tico medio de estimaci3n y ra3z del error cuadr3tico medio de predicci3n suponiendo que el retardo es desconocido.

M	$T = 30$	BIC	BIC*	AIC	AIC*	AICc	AICc*	AICu	AICu*	C_1	C_1^*	Cc	Cc*	Cu	Cu*
7	(p_1, p_2)	205	303	163	248	739	808	862	897	292	403	765	824	863	913
7	d	540	559	553	571	582	616	589	617	540	540	573	578	573	575
7	(p_1, p_2, d)	135	187	117	165	463	517	527	566	187	240	469	499	517	536
7	$RMSE$	0.65	0.64	0.63	0.62	0.43	0.42	0.42	0.41	0.55	0.55	0.41	0.41	0.41	0.41
7	$RMSPE$	1.62	1.61	1.64	1.62	1.18	1.17	1.17	1.17	1.29	1.28	1.19	1.19	1.18	1.18
8	(p_1, p_2)	184	249	147	221	733	802	846	879	292	390	754	809	860	885
8	d	569	555	562	568	601	590	608	598	611	599	649	632	653	633
8	(p_1, p_2, d)	119	146	98	141	461	485	527	538	203	254	512	526	577	570
8	$RMSE$	0.90	0.90	0.88	0.88	0.68	0.69	0.67	0.67	0.78	0.78	0.67	0.68	0.66	0.67
8	$RMSPE$	1.90	1.86	1.90	1.86	1.27	1.25	1.24	1.22	1.52	1.51	1.28	1.27	1.25	1.24
M	$T = 50$	BIC	BIC*	AIC	AIC*	AICc	AICc*	AICu	AICu*	C_1	C_1^*	Cc	Cc*	Cu	Cu*
7	(p_1, p_2)	223	340	85	168	529	624	744	800	187	292	554	645	746	812
7	d	335	337	337	320	364	365	369	371	384	380	394	403	406	419
7	(p_1, p_2, d)	99	130	40	63	225	250	296	318	107	149	252	288	326	360
7	$RMSE$	0.63	0.61	0.61	0.60	0.48	0.47	0.47	0.47	0.53	0.53	0.46	0.46	0.45	0.45
7	$RMSPE$	1.25	1.22	1.30	1.26	1.17	1.16	1.14	1.14	1.24	1.23	1.18	1.18	1.16	1.15
8	(p_1, p_2)	247	325	105	183	570	655	773	824	217	293	594	668	784	834
8	d	419	421	400	423	483	474	500	483	480	478	528	521	544	529
8	(p_1, p_2, d)	146	182	66	115	325	347	419	427	143	178	359	380	455	460
8	$RMSE$	0.84	0.84	0.82	0.82	0.68	0.68	0.68	0.68	0.72	0.72	0.66	0.66	0.65	0.66
8	$RMSPE$	1.44	1.44	1.39	1.35	1.20	1.20	1.17	1.16	1.31	1.29	1.20	1.19	1.18	1.18
M	$T = 100$	BIC	BIC*	AIC	AIC*	AICc	AICc*	AICu	AICu*	C_1	C_1^*	Cc	Cc*	Cu	Cu*
7	(p_1, p_2)	652	747	221	330	388	480	667	773	235	316	401	481	662	773
7	d	491	542	489	527	504	542	522	527	552	567	567	572	582	582
7	(p_1, p_2, d)	351	421	135	192	210	286	376	436	195	251	301	351	436	481
7	$RMSE$	0.53	0.52	0.52	0.50	0.49	0.46	0.49	0.47	0.49	0.49	0.48	0.47	0.49	0.46
7	$MSPE$	1.04	1.03	1.05	1.05	1.06	1.06	1.04	1.02	1.06	1.06	1.04	1.02	1.04	1.03
8	(p_1, p_2)	808	888	371	466	547	632	838	863	421	532	602	662	798	863
8	d	732	727	667	667	677	692	717	712	747	773	773	788	788	793
8	(p_1, p_2, d)	632	662	291	371	431	486	632	637	376	456	517	562	657	697
8	$RMSE$	0.58	0.59	0.60	0.59	0.56	0.57	0.56	0.56	0.56	0.56	0.54	0.57	0.55	0.56
8	$RMSPE$	1.04	1.03	1.05	1.05	1.04	1.04	1.04	1.03	1.06	1.06	1.07	1.05	1.05	1.05

estacionarios TAR,

$$(M7) \begin{cases} x_t = -0.8x_{t-1} + a_{1t}, & x_{t-1} \leq 0 \\ x_t = -0.2x_{t-1} + a_{2t}, & x_{t-1} > 0 \end{cases} \quad (M8) \begin{cases} x_t = 0.5x_{t-1} + a_{1t}, & x_{t-1} \leq 0 \\ x_t = -0.5x_{t-1} + a_{2t}, & x_{t-1} > 0 \end{cases}$$

donde $a_{jt} \sim N(0, 1)$, $j = 1, 2$. Basado en la sección 5.4, se compara el comportamiento de los criterios en (5.20), (5.24) con respecto a los criterios en (5.26) y los criterios en (5.21), (5.23) y (5.25) con respecto a los criterios en (5.27). En todos los casos, se generan 1000 series de los modelos M7 y M8 con tamaños muestrales $T = 31, 51$ y 101 . Se procede como en Wong y Li (1998) y De Gooijer (2001) utilizando una malla para estimar el parámetro umbral, r . Se ajusta cada modelo a las primeras $T - 1$ observaciones de cada serie mediante máxima verosimilitud condicional en cada modelo y se obtiene el término de corrección en (5.16) en cada regimen. En primer lugar, se supone que el parámetro retardo es conocido y se fija $p_1^{\max} = p_2^{\max} = 5$ para $T = 31, 51$ y 101 , por lo que teniendo en cuenta que el número de posibles valores del parámetro umbral es $(T - 1)/2$, se comparan 375, 625 y 1250 modelos respectivamente. En todo caso, se consideran las siguientes medidas del comportamiento de los criterios de selección: (a) la frecuencia de detección del orden correcto $(p_1, p_2) = (1, 1)$, (b) la raíz del error cuadrático medio de estimación del parámetro umbral y (c) la raíz del error de predicción mínimo cuadrático para la última observación basado en el modelo seleccionado por cada criterio, los parámetros ajustados y el verdadero valor. Los resultados aparecen en la Tabla 5.4. Se puede observar que para tamaños muestrales pequeños, $T = 30$, la mejora en el número de veces en el se selecciona el modelo correcto puede ser tan grande como 30.5 % (ver C_1 y C_1^* en M7), para $T = 50$ puede alcanzar el 50.5 % (ver AIC y AIC* en M7) y para $T = 100$ puede alcanzar el 22.6 % (ver AIC y AIC* en M7). Hay que notar que los criterios AICu, AICu*, Cu y Cu* tienen mayores frecuencias de detección para $T = 30$ pero la frecuencia de detección disminuye cuando el tamaño muestral crece.

Por otro lado, la raíz del error medio cuadrático de estimación (RMSE) del parámetro umbral es muy cercana para los criterios originales y modificados, mientras que el error medio cuadrático de predicción (RMSPE) es, en general, menor para los criterios corregidos.

A continuación, se supone que el retardo es desconocido y se fija $p_1^{\max} = p_2^{\max} = 5$ y $d^{\max} = 4$ para $T = 31, 51$ and 101 , por lo que teniendo en cuenta que el número de posibles valores del umbral es $(T - 1)/2$, se comparan 1500, 2500 y 5000 modelos respectivamente. En todo caso, se consideran las siguientes medidas del comportamiento de los criterios: (a) la frecuencia de detección de los ordenes correctos, $(p_1, p_2) = (1, 1)$, (b) la frecuencia de detección de seleccionar el retardo correcto $d = 1$, (c) la frecuencia de detección de los ordenes y retardo correcto, (d) la raíz del error mínimo cuadrático de estimación del umbral, y (e) la raíz del error de predicción mínimo cuadrático para la última observación basada en el modelo seleccionado por cada criterio, los parámetros ajustados y el verdadero valor. Los resultados aparecen en la Tabla 5.5. Se puede observar que para tamaño muestral pequeño, $T = 30$, la mejora en el número de veces en que los ordenes correctos $(p_1, p_2, d) = (1, 1, 1)$ son seleccionados puede ser tan grande como el 43.8 % (ver AIC y AIC* en M8), para $T = 50$ puede alcanzar el 74.2 % (ver AIC y AIC* en M8) y para $T = 100$ puede alcanzar el 42.2 %

(ver AIC y AIC* en M7). Como en el caso en que d se supone conocido, los criterios AICu, AICu*, Cu y Cu* tienen una frecuencia de detección mayor para los ordenes autorregresivos y el retardo para $T = 30$ pero la frecuencia de detección decrece cuando el tamaño muestral se incrementa. Se nota que a veces los criterios corregidos tienen una frecuencia de detección menor del parámetro retardo pero esto no es un problema debido a que el interés en este caso es seleccionar correctamente los ordenes autorregresivos y el retardo al mismo tiempo. Respecto a los valores de RMSE y el RMSPE, los resultados son similares al caso en que d es conocido.

5.6. Conclusiones

En este capítulo, se ha demostrado que analizar selección de modelos en modelos ARMA como un problema de discriminación cuadrática proporciona un enfoque unificador para derivar criterios de selección de modelos. También este enfoque sugiere una diferente definición de esperanza de la verosimilitud a la propuesta por Akaike (1973). Este enfoque lleva a la inclusión de un término de corrección en los criterios que no modifican su comportamiento en muestras grandes pero que hacen mejorar el comportamiento en muestras pequeñas. Por lo tanto, se ha propuesto una familia de criterios que generalizan los criterios de selección comunmente utilizados en la práctica. Estas ideas son extendidas al caso de modelos autorregresivos por umbrales (TAR) y se generaliza el enfoque propuesto para estos modelos de series no lineales. Un estudio de Monte Carlo demuestra que esta familia mejora el comportamiento en muestras finitas de criterios tales como el AIC, AICc y BIC, para modelos ARMA, y AIC, AICc, BIC y algunos criterios de validación cruzada para modelos TAR. En particular, para tamaños muestrales pequeños y medios, la frecuencia de selección del verdadero modelo aumenta para los criterios consistentes y la raíz del error de predicción cuadrático medio disminuye para los criterios eficientes. Estos resultados son obtenidos para modelos lineales ARMA y modelos TAR en los que se asume que el umbral y los parámetros son desconocidos.

5.7. Apéndice

Demostración del Lema 5.1. Utilizando (5.2), se tiene que,

$$E_{\alpha_0} [\log p(y | \hat{\alpha}_{p,q})] = -\frac{T}{2} \log 2\pi - \frac{T}{2} \log \hat{\sigma}_{p,q}^2 - \frac{1}{2} \log |Q_T(\hat{\beta}_{p,q})| - \frac{1}{2} E_{\alpha_0} \left[\frac{S_y(\hat{\beta}_{p,q})}{\hat{\sigma}_{p,q}^2} \right],$$

donde $S_y(\hat{\beta}_{p,q}) = y' Q_T^{-1}(\hat{\beta}_{p,q}) y$. Suponiendo que el modelo $M_{p,q}$ es el verdadero modelo, Brockwell y Davies (1991) demuestran que,

$$E \left[\frac{S_y(\hat{\beta}_{p,q})}{\hat{\sigma}_{p,q}^2} \right] \simeq \frac{E[S_y(\hat{\beta}_{p,q})]}{E[\hat{\sigma}_{p,q}^2]} = \frac{T(T+p+q)}{(T-p-q-2)} + O_p(1), \quad (5.28)$$

lo que conduce a (5.6).

Por otro lado, utilizando que $T \log(1 - \frac{(p+q)}{T}) = -(p+q) + o(1)$, se tiene que,

$$\begin{aligned} & T \log 2\pi + T \log \frac{T}{T-(p+q)} \hat{\sigma}_{p,q}^2 + \log |Q_T(\hat{\beta}_{p,q})| = \\ & = T \log 2\pi - T \log \left(1 - \frac{(p+q)}{T} \right) + T \log \hat{\sigma}_{p,q}^2 + \log |Q_T(\hat{\beta}_{p,q})| = \\ & = T \log 2\pi + T \log \hat{\sigma}_{p,q}^2 + (p+q) + \log |Q_T(\hat{\beta}_{p,q})| + o_p(1). \end{aligned}$$

Más aun, de (5.28),

$$E \left[\frac{S_x(\hat{\beta}_{p,q})}{\frac{T}{T-(p+q)} \hat{\sigma}_{p,q}^2} \right] = \frac{(T+p+q)T}{\frac{T}{T-(p+q)}(T-p-q-2)} + O_p(1) = (T+p+q) + O_p(1),$$

lo que prueba (5.6). ■

Demostración del Lema 5.2. Sea,

$$h(\alpha_{p,q}) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_T(\alpha_{p,q})| - \frac{1}{2} x' \Sigma_T(\alpha_{p,q})^{-1} x + \log p(\alpha_{p,q} | M_{p,q}),$$

entonces, aplicando el método de Laplace, ver Tierney y Kadane (1986), se obtiene,

$$p(x | M_{p,q}) \approx (2\pi)^{\frac{p+q+1-T}{2}} |H(\hat{\alpha}_{p,q})|^{\frac{1}{2}} |\Sigma_T(\hat{\alpha}_{p,q})|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} x' \Sigma_T(\hat{\alpha}_{p,q})^{-1} x \right) p(\hat{\alpha}_{p,q} | M_{p,q}),$$

donde $\hat{\alpha}_{p,q}$ es el estimador máximo verosímil de $\alpha_{p,q}$ y H es menos el Hessiano inverso de h evaluado en $\hat{\alpha}_{p,q}$. Entonces,

$$\begin{aligned} \log p(x | M_{p,q}) & \approx \frac{p+q+1-T}{2} \log(2\pi) + \frac{1}{2} \log |H(\hat{\alpha}_{p,q})| - \\ & - \frac{1}{2} \log |\Sigma_T(\hat{\alpha}_{p,q})| - \frac{1}{2} x' \Sigma_T(\hat{\alpha}_{p,q})^{-1} x + \log p(\hat{\alpha}_{p,q} | M_{p,q}). \end{aligned}$$

Raftery, Madigan y Volinsky (1995) demuestran que $\log |H(\hat{\alpha}_{p,q})| = -(p+q+1) \log T + O_p(1)$ porque $H(\hat{\alpha}_{p,q})$ es asintóticamente igual a la matriz de información observada, que es asintóticamente igual a T veces una matriz constante. Entonces,

$$\begin{aligned} \log p(x|M_{p,q}) &= \frac{1}{2} (p+q+1-T) \log(2\pi) - \frac{1}{2} (p+q+1) \log T - \frac{1}{2} \log |\Sigma_T(\hat{\alpha}_{p,q})| \\ &- \frac{1}{2} x' \Sigma_T(\hat{\alpha}_{p,q})^{-1} x + \log p(\hat{\alpha}_{p,q}|M_{p,q}) + O(1) = \frac{1}{2} (p+q+1-T) \log(2\pi) - \frac{1}{2} (p+q+1) \log T \\ &- \frac{T}{2} \log \hat{\sigma}_{p,q}^2 - \frac{1}{2} \log |Q_T(\hat{\beta}_{p,q})| - \frac{1}{2} T + \log p(\hat{\alpha}_{p,q}|M_{p,q}) + O_p(1), \end{aligned}$$

que demuestra el resultado. ■

Demostración del Teorema 5.3. Shibata (1980) considera criterios de selección de la forma:

$$S_T^o(p) = (T - p_{\max} + \delta_T(p) + 2p) \hat{\sigma}_p^2.$$

El orden seleccionado por los criterios $S_T^o(p)$ es eficiente si $\delta_T(p)$ verifica las condiciones impuestas en el Teorema 4.2 de Shibata:

1. $\text{plím}_{T \rightarrow \infty} \left(\max_{1 \leq p \leq p_{\max}} \frac{|\delta_T(p)|}{T - p_{\max}} \right) = 0,$
2. $\text{plím}_{T \rightarrow \infty} \left(\max_{1 \leq p \leq p_{\max}} \frac{|\delta_T(p) - \delta_T(p_T^*)|}{(T - p_{\max}) L_T(p)} \right) = 0,$

donde plím denota limite en probabilidad, $L_T(p)$, es la siguiente función,

$$L_T(p) = \frac{p\sigma_a^2}{T - p_{\max}} + \sum_{i=p+1}^{\infty} \sum_{j=p+1}^{\infty} \phi_i \phi_j \Sigma_{ij}$$

donde $\Sigma_{ij} = \text{Cov}(x_t, x_{t-|i-j|})$ y p_T^* es una secuencia de enteros positivos con $1 \leq p_T^* \leq p_{\max}$ que alcanzan el mínimo de $L_T(p)$ para cada T (ver Shibata, 1980, p.154). El criterio AIC puede escribirse en términos de $S_T^o(p)$ tomando $\delta_T(p) = \delta_T^{AIC}(p) = T \exp\left(\frac{2p}{T}\right) - (T - p_{\max}) - 2p$. Shibata (1980) demuestra que este término verifica las dos condiciones, y esto proporciona la eficiencia asintótica del AIC. Podemos escribir AIC* en términos de $S_T^o(p)$ tomando $\delta_T(p) = \delta_T^{AIC^*}(p) = T \exp\left(\frac{2p}{T}\right) \left(\log |Q_T(\hat{\beta}_p)|\right)^{\frac{1}{T}} - (T - p_{\max}) - 2p$. Por lo tanto,

$$\delta_T^{AIC^*}(p) = \delta_T^{AIC}(p) - T \exp\left(\frac{2p}{T}\right) \left(1 - \left(\log |Q_T(\hat{\beta}_p)|\right)^{\frac{1}{T}}\right).$$

A continuación, se demuestra que $\delta_T^{AIC^*}(p)$ verifica ambas condiciones. Primero, se escribe,

$$\frac{|\delta_T^{AIC^*}(p)|}{T - p_{\max}} = \left| \frac{\exp\left(\frac{2p}{T}\right) \left(\log |Q_T(\hat{\beta}_p)|\right)^{\frac{1}{T}}}{1 - \frac{p_{\max}}{T}} - \frac{\frac{2p}{T}}{1 - \frac{p_{\max}}{T}} - 1 \right|. \quad (5.29)$$

Hannan (1973) demuestra que $(\log |Q_T(\gamma)|)^{\frac{1}{T}} \rightarrow 1$, para cada γ perteneciente al espacio paramétrico, y consecuentemente, $(\log |Q_T(\hat{\beta}_p)|)^{\frac{1}{T}} \rightarrow 1$ y el límite cuando $T \rightarrow \infty$ del máximo de los valores (5.29) en el conjunto $1 \leq p \leq p_{\max}$ es 0. Esto prueba la primera condición.

Para la segunda condición, se escribe la siguiente descomposición:

$$\begin{aligned} & \frac{|\delta_T^{AIC^*}(p) - \delta_T^{AIC^*}(p_T^*)|}{(T - p_{\max}) L_T(p)} \leq \frac{|\delta_T^{AIC}(p) - \delta_T^{AIC}(p_T^*)|}{(T - p_{\max}) L_T(p)} + \\ & + \frac{\left| T \exp\left(\frac{2p_T^*}{T}\right) \left(1 - \left(\log |Q_T(\hat{\beta}_{p_T^*})|\right)^{\frac{1}{T}}\right) - T \exp\left(\frac{2p}{T}\right) \left(1 - \left(\log |Q_T(\hat{\beta}_p)|\right)^{\frac{1}{T}}\right) \right|}{(T - p_{\max}) L_T(p)}. \end{aligned}$$

Shibata (1980) demuestra que el primer término tiende a 0 implicando que AIC es eficiente. Para la segunda expresión, para cada p tal que $1 \leq p \leq p_{\max}$ incluyendo p_T^* , se puede demostrar que,

$$\lim_{T \rightarrow \infty} T \exp\left(\frac{2p}{T}\right) \left(1 - \left(\log |Q_T(\hat{\beta}_p)|\right)^{\frac{1}{T}}\right) = -\log \left(-\sum_{i=1}^p i \log(1 - \phi_{ii}^2(\beta_p)) \right) < \infty.$$

Como el límite es acotado para cada p y $(T - p_{\max}) L_T(p) \rightarrow \infty$ cuando $T \rightarrow \infty$, para cada $1 \leq p \leq p_{\max}$, la segunda expresión también tiende a 0. Entonces, $\delta_T^{AIC^*}(p)$ verifica la segunda condición. Por lo tanto, AIC* es eficiente y como AICc* es asintóticamente equivalente a AIC*, AICc* es también eficiente. ■

Capítulo 6

Futuras líneas de investigación

Es este capítulo se describen algunas futuras líneas de investigación que extienden el trabajo realizado en esta Tesis.

En los capítulos 3 y 4 se han propuesto métodos de detección de atípicos y cambios de nivel y cambios de varianzas y de correlaciones en series multivariantes por separado. Parece razonable desarrollar un método que detecte todos estos cambios de una vez. La primera línea de extensión es construir un procedimiento para identificar simultáneamente todos estos efectos y desarrollar procedimientos que posibiliten su tratamiento conjunto. Se estudiará la capacidad de un modelo que incluya todos los efectos mencionados para ajustar series multivariantes. Algunos puntos a estudiar son los siguientes. En primer lugar, analizar la influencia del grado de diferenciación de la serie en los atípicos. Por ejemplo, un atípico innovativo tiene un efecto transitorio en una serie estacionaria y un efecto permanente muy parecido a un cambio de nivel en una serie $I(1)$. Por ello, es importante estudiar cual es la mejor manera de modelizar estos efectos para evitar posibles confusiones. En segundo lugar, comparar el modelo que incluya todos los efectos con otros tales como los de heterocedasticidad condicional, para determinar si ofrecen un mejor ajuste para algunas series. La razón de esta comparación es que se han estimado algunas series financieras suponiendo cambios discretos en la varianza y distribución Gaussiana en el término de error, y se ha comprobado como se reduce drásticamente la correlación entre los cuadrados de los residuos obteniendo un ajuste adecuado. En tercer lugar, analizar modelos de heterocedasticidad condicional para series multivariantes que incluyan atípicos y cambios en las varianzas marginales. Como se ha mencionado en el capítulo 2, algunos artículos han tratado estos modelos en el caso univariante pero no en el multivariante.

La segunda línea de extensión es utilizar los criterios de selección de modelos para detectar atípicos y cambios de varianza. Por ejemplo, Kitagawa (1979) utiliza el AIC para detectar atípicos en observaciones multivariantes y Chen y Gupta (1997) utilizan el BIC para detectar cambios de varianza en series de ruido blanco. Una segunda línea de investigación es obtener métodos de detección de atípicos y cambios de varianza mediante el uso de los criterios de

selección propuestos en el capítulo 5. Supongamos una serie univariante que sigue un modelo ARMA y se quiere contrastar la presencia de un atípico o cambio de varianza en $t = h$. La idea es comparar el valor del criterio de selección de un modelo que incluya el atípico o el cambio de varianza como un parámetro a estimar, con el valor del criterio para un modelo que no lo incluya. Si el valor del criterio es menor para el modelo con el atípico o el cambio, entonces se asume que la serie tiene uno de estos efectos en el instante $t = h$. La ventaja de este método con respecto a uno basado en contrastes es que se obtiene una regla objetiva para determinar la existencia de un atípico o de un cambio de varianza, que es independiente de niveles de significatividad. Los criterios BIC y BIC* parecen adecuados para detectar atípicos, ya se ha demostrado que asintóticamente siempre seleccionan el modelo adecuado. También se ha demostrado que los criterios eficientes como el AIC, AIC*, AICc y AICc*, no parecen adecuados en este sentido ya que incluso asintóticamente existe una probabilidad positiva de seleccionar el modelo con el atípico aunque no exista. Se estudiará la aplicación de este tipo de procedimientos en series univariantes y multivariantes y en modelos lineales y no lineales.

La tercera extensión es explorar la posibilidad de detectar cambios de las varianzas y de las correlaciones en series multivariantes mediante el uso de proyecciones. Se analizará la proyección del vector de series y del vector de residuos. En ambos casos, el primer paso debe ser obtener los efectos de los cambios en las series proyectadas. Para el caso de los residuos, si suponemos el momento de cambio conocido, una posibilidad es obtener la dirección que maximice o minimice el cociente entre varianzas de los residuos proyectados antes y después del momento de cambio. Si Σ y Ω denotan estas matrices, entonces, las direcciones buscadas son las de los vectores asociados a los autovalores máximo y mínimo de la matriz $\Sigma^{-1}\Omega$. Un inconveniente es que en la serie univariante proyectada se pierde la información referente a si el cambio es de varianzas, de correlaciones o de ambos, por lo que habría que realizar una comparación posterior a detectar un cambio. Algunos puntos a investigar serán también como evitar detectar cambios espúreos y como obtener una dirección que sea independiente del momento de cambio.

Una cuarta línea de investigación es construir un procedimiento para identificar cambios estructurales basado en estadísticos multivariantes, como los que aparecen en el capítulo 4, que tengan en cuenta posibles cambios en los parámetros. También se analizará el uso de proyecciones para detectar estos cambios obtenidos para series proyectadas, debido a que un cambio en los parámetros también afecta al modelo de la serie proyectada. Para ello es necesario, como en los casos anteriores, determinar un índice de proyección adecuado.

Una quinta línea de investigación es ampliar los resultados de selección de modelos del capítulo 5 al caso de series multivariantes. El planteamiento del problema mediante análisis discriminante para modelos ARMA parece que se puede extender de manera natural al caso multivariante. Un punto a investigar es si de este planteamiento se puede derivar un término de corrección similar al caso univariante y si dicho término tiene en cuenta las correlaciones de las componentes del vector de series.

Bibliografía

- Abraham, B y Box, G. E. P. (1979) "Bayesian Analysis of Some Outliers Problems in Time Series", *Biometrika*, **66**, 229-236.
- Abraham, B. y Wei, W. W. S. (1984) "Inferences About the Parameters of a Time Series Model With Changing Variance", *Metrika*, **31**, 183-194.
- Akaike, H. (1969) "Fitting Autoregressive Models for Prediction", *Annals of the Institute of Statistical Mathematics*, **21**, 243-247.
- Akaike, H. (1973) "Information Theory and an Extension of the Maximum Likelihood Principle", Proceedings of the 2nd International Symposium on Information Theory (Akademiai Kiadó, Budapest), 267-281.
- Akaike, H. (1974) "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, **6**, 716-723.
- Akaike, H. (1977) "On Entropy Maximization Principle", Applications of Statistics, North-Holland: Amsterdam, 27-41.
- Akaike, H. (1979) "A Bayesian Extension of the Minimum AIC Procedure of Autoregressive Model Fitting", *Biometrika*, **66**, 237-242.
- Akaike, H. (1981) "Likelihood of a Model and Information Criteria", *Journal of Econometrics*, **16**, 3-14.
- Ali, M. M. y Giaccotto, C. (1982) "The Identical Distribution Hypothesis for Stock Market Prices – Location- And Scale-Shift Alternatives", *Journal of the American Statistical Association*, **77**, 19-28.
- Bai, J. (1994) "Least Squares Estimation of a Shift in Linear Processes", *Journal of Time Series Analysis*, **15**, 453-472.
- Balke, N. S. (1993) "Detecting Level Shifts in Time Series", *Journal of Business and Economic Statistics*, **11**, 81-92.
- Barnett, V. y Lewis, T. (1994) *Outliers in Statistical Data*, 3rd Ed. John Wiley & Sons: Chichester.

- Baufays, P. y Rasson, J. P. (1985) "Variance Changes in Autoregressive Models", *Time Series: Theory and Methods*, 2nd Ed. Springer: New York, 119-127.
- Bhansali, R. y Downham, D. Y. (1977) "Some Properties of the Order of an Autoregressive Model Selected by a Generalization of Akaike's FPE Criterion", *Biometrika*, **64**, 547-552.
- Bhansali, R. (1986) "The Criterion Autoregressive Transfer Function of Parzen", *Journal of Time Series Analysis*, **7**, 79-104.
- Bianco, A. M., García Ben, M., Martínez, E. J. y Yohai, V. J. (2001) "Outlier Detection in Regression Models with ARIMA Errors using Robust estimates", *Journal of Forecasting*, **20**, 565-579.
- Billingsley, P. (1968) *Convergence of Probability Measures*. John Wiley & Sons.
- Bollerslev, T. (1986) "Generalized Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics*, **31**, 307-327.
- Box, G. E. P. y Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control*. Holden-day: Oakland.
- Box, G. E. P. y Tiao, G. C. (1965) "A Change in Level of a Non-stationary Time Series", *Biometrika*, **52**, 181-192.
- Box, G. E. P. y Tiao, G. C. (1975) "Intervention Analysis With Applications to Economic and Environmental Problems", *Journal of the American Statistical Association*, **70**, 70-79.
- Bozdogan, H. (1987) "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions", *Psychometrika*, **52**, 245-370.
- Brockwell, P. J. y Davis, R. A. (1991) *Time Series: Theory and Methods*. Springer-Verlag: New York.
- Bruce, A. G. y Martin, R. D. (1989) "Leave-k-out Diagnostics for Time Series", *Journal of the Royal Statistical Society B*, **51**, 363-401.
- Burnham, K. P. y Anderson, D. R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer: New York.
- Buseti, F. y Taylor, A. M. R. (2003) "Variance Shifts, Structural Breaks and Stationarity Tests", *Journal of Business and Economic Statistics*, **21**, 510-531.
- Bustos, O. H. (1982) "General M-estimates for Contaminated pth-order Autoregressive Processes: Consistency and Asymptotic Normality", *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **59**, 491-504.

- Bustos, O. H. y Yohai, V. J. (1986) "Robust Estimates for ARMA Models", *Journal of the American Statistical Association*, **81**, 155-168.
- Carnero, M. A., Peña, D. y Ruiz, E. (2001) "Outliers and Conditional Autoregressive Heteroscedasticity in Time Series", *Estadística*, **53**, 143-213.
- Carnero, M. A., Peña, D. y Ruiz, E. (2003) "Detecting Level Shifts in the Presence of Conditional Heteroscedasticity", Working Paper, Universidad Carlos III de Madrid.
- Cavanaugh, J. E. (1997) "Unifying the Derivations for the Akaike and Corrected Akaike Information Criteria", *Statistics and Probability Letters*, **33**, 201-208.
- Cavanaugh, J. E. y Neath, A. A. (1999) "Generalizing the Derivation of the Schwarz Information Criterion", *Communications in Statistics: Theory and Methods*, **28**, 49-66.
- Cavanaugh, J. E. y Shumway, R. H. (1997) "A Bootstrap Variant of AIC for State-Space Model Selection", *Statistica Sinica*, **7**, 473-496.
- Chan, K. S. (1993) "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model", *Annals of Statistics*, **21**, 520-533.
- Chan W. (1995) "Understanding the Effect of Time Series Outliers in Sample Autocorrelations", *Test*, **4**, 179-186.
- Chang I. y Tiao G. C. (1983) "Estimation of Time Series Parameters in the Presence of Outliers", Technical Report, Statistics Research Center, University of Chicago.
- Chang I., Tiao, G. C. y Chen, C. (1988) "Estimation of Time Series Parameters in the Presence of Outliers", *Technometrics*, **3**, 193-204.
- Chen, J. y Gupta, A. K. (1997) "Testing and Locating Variance Change-points with Application to Stock Prices", *Journal of the American Statistical Association*, **92**, 739-747.
- Chen, C. y Liu, L. (1993, a) "Joint Estimation of Model Parameters and Outlier Effects in Time Series", *Journal of the American Statistical Association*, **88**, 284-297.
- Chen, C. y Liu, L. (1993, b) "Forecasting Time Series with Outliers", *Journal of Forecasting*, **12**, 13-35.
- Chen, C. y Tiao, G. C. (1990) "Random Level-Shift Time Series Models, ARIMA Approximations, and Level-Shift Detection", *Journal of the Business and Economic Statistics*, **8**, 83-97.
- Cheng, B. y Tong, H. (1992) "On Consistent Nonparametric Order Determination and Chaos", *Journal of the Royal Statistical Society B*, **54**, 427-449.
- Chernick, M. R., Downing, D. D. y Pike, D. H. (1982) "Detecting Outliers in Time Series Data", *Journal of the American Statistical Association*, **77**, 743-747.

- Chipman, H., George, E. I. y McCulloch, R. E. (2001) "The Practical Implementation of Bayesian Model Selection", *Model Selection*, Institute of Mathematical Statistics: Ohio, 65-134.
- Chow, G. C. (1981) "A Comparison of the Information and Posterior Probability Criteria for Model Selection", *Journal of Econometrics*, **16**, 21-33.
- Davis, W. W. (1979) "Robust Methods for Detection of Shifts of the Innovation Variance of a Time Series", *Technometrics*, **21**, 313-320.
- De Gooijer, J. G. (2001) "Cross-validation Criteria for SETAR Model Selection", *Journal of Time Series Analysis*, **22**, 267-281.
- Denby, L. y Martin, D. (1979) "Robust Estimation of the First Order Autoregressive Parameter", *Journal of the American Statistical Association*, **74**, 140-146.
- Deutsch, S. J., Richards, J. E. y Swain, J. J. (1990) "Effects of a Single Outlier on ARMA Identification", *Communications in Statistics: Theory and Methods*, **19**, 2207-2227.
- Durbin, J. (1960) "The Fitting of Time Series Model", *Review of the International Institute of Statistics*, **28**, 233-244.
- Efron, B. (1983) "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-validation", *Journal of the American Statistical Association*, **78**, 316-331.
- Efron, B. (1986) "How Biased is the Apparent Error Rate of a Prediction Rule?", *Journal of the American Statistical Association*, **81**, 461-470.
- Efron, B. y Tibshirani, R. J. (1993) *An Introduction to the Bootstrap*. Chapman-Hall: London.
- Engle, R. (1982) "Autoregressive Conditional Heterocedastic Models With Estimates of the Variance of United Kingdom Inflation", *Econometrica*, **50**, 987-1007.
- Engle, R. y Granger, C. W. J. (1987) "Co-integration and Error Correction: Representation, Estimation and Testing", *Econometrica*, **55**, 251-267.
- Findley, D. F. (1985) "On the Unbiasedness Property of AIC for Exact or Approximating Linear Stochastic Time Series Models", *Journal of Time Series Analysis*, **6**, 229-252.
- Fox, A. J. (1972) "Outliers in Time Series", *Journal of the Royal Statistical Society B*, **34**, 350-363.
- Galeano, P. y Peña, D. (2000) "Multivariate Analysis in Vector Time Series", *Resenhas*, **4**, 383-403.
- Galeano, P. y Peña, D. (2004, a) "Variance Changes Detection in Multivariate Time Series", Working Paper, Universidad Carlos III de Madrid.

- Galeano, P. y Peña, D. (2004, b) "Model Selection Criteria and Quadratic Discrimination in ARMA and SETAR Time Series Models", Working Paper, Universidad Carlos III de Madrid.
- Galeano, P., Peña, D. y Tsay, R. S. (2004) "Outlier Detection in Multivariate Time Series via Projection Pursuit", Manuscrito.
- García Ben, M., Martínez, E. J. y Yohai, V. J. (1999) "Robust Estimation in Vector Autoregressive Moving-Average Models", *Journal of Time Series Analysis*, **20**, 381-399.
- Grubb, H. (1992) "A Multivariate Time Series Analysis of Some Flour Price Data", *Applied Statistics*, **26**, 279-284.
- Hamori, S. (1997) "Testing for a Unit Root in the Presence of a Variance Shift", *Economics Letters*, **57**, 245-253.
- Hannan, E. J. (1973) "The Asymptotic Theory of Linear Time-Series Models", *Journal of Applied Probability* **10**, 130-145.
- Hannan, E. J. (1980) "Estimation of the Order of an ARMA Process", *Annals of Statistics*, **8**, 1071-1081.
- Hannan, E. J. y Deistler, M. (1988) *The Statistical Theory of Linear Systems*. John Wiley: New York.
- Hannan, E. J. y Kavalieris, L. (1984) "A Method for Autoregressive-Moving Average Estimation", *Biometrika*, **72**, 273-280.
- Hannan, E. J. y Quinn, B. G. (1979) "The Determination of the Order of an Autoregression", *Journal of the Royal Statistical Society B*, **41**, 190-195.
- Hannan, E. J. y Rissanen, J. (1982) "Recursive Estimation of Mixed Autoregressive-Moving Average Order", *Biometrika*, **69**, 81-94.
- Harvey, A. (1989) *Forecasting Structural Time Series Models and the Kalman Filter*, 2nd edn. Cambridge University Press: Cambridge.
- Harvey, A., Ruiz, E. y Shephard, N. (1994) "Multivariate Stochastic Variance Models", *Review of Economic Studies*, **61**, 247-264.
- Haughton, D. M. A. (1988) "On the Choice of a Model to Fit Data From an Exponential Family", *Annals of Statistics*, **16**, 342-355.
- Hillmer, S. (1984) "Monitoring and Adjusting Forecasts in the Presence of Additive Outliers", *Journal of Forecasting*, **3**, 205-215.
- Hotta, L. K. (1993) "The Effect of Additive Outliers on the Estimates From Aggregated and Dissaggregated ARIMA Models", *International Journal of Forecasting*, **9**, 85-93.

- Hsu, D. A. (1977) "Tests for Variance Shift at an Unknown Time Point", *Applied Statistics*, **26**, 279-284.
- Hsu, D. A. (1979) "Detecting Shifts of Parameter in Gamma Sequences With Applications to Stock Price and Air Traffic Flow Analysis", *Journal of the American Statistical Association*, **74**, 31-40.
- Hsu, D. A., Miller, R. B. y Wichern, D. W. (1974) "On the Stable Paretian Behavior of Stock-Market Prices", *Journal of the American Statistical Association*, **69**, 108-113.
- Huber, P. (1973) "Robust Regression: Asymptotics, Conjectures and Monte Carlo", *Annals of Statistics*, **1**, 799-821.
- Huber, P. (1985) "Projection Pursuit (with discussion)", *Annals of Statistics*, **13**, 435-525.
- Hurvich, C. M., Shumway, R. y Tsai, C. (1990) "Improved Estimators of Kullback-Leibler Information for Autoregressive Model Selection in Small Samples", *Biometrika*, **77**, 709-719.
- Hurvich, C. M. y Tsai, C. (1989) "Regression and Time Series Model Selection in Small Samples", *Biometrika*, **76**, 297-307.
- Hurvich, C. M. y Tsai, C. (1991) "Bias of the Corrected AIC Criterion for Underfitted Regression and Time Series Models", *Biometrika*, **78**, 499-509.
- Hurvich, C. M. y Tsai, C. (1993) "A Corrected Akaike Information Criterion for Vector Autoregressive Model", *Journal of Time Series Analysis*, **14**, 271-279.
- Inclán, C. (1993) "Detection of Multiple Tests of Variances Using Posterior Odds", *Journal of Business and Economic Statistics*, **11**, 189-200.
- Inclán, C. y Tiao, G. C. (1994) "Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance", *Journal of the American Statistical Association*, **89**, 913-923.
- Ishiguro, M., Sakamoto, Y. y Kitagawa, G. (1997) "Bootstrapping Log Likelihood and EIC, an Extension of AIC", *Annals of the Institute of Statistical Mathematics*, **49**, 411-434.
- Jones, M. C. y Sibson, R. (1987) "What is Projection Pursuit (with discussion)?", *Journal of the Royal Statistical Society A*, **150**, 1-36.
- Justel, A., Peña, D. y Sánchez, M. J. (1993) "Grupos de Atípicos en Modelos Económicos", *Cuadernos Económicos del ICE*, **55**, 285-325.
- Justel, A., Peña, D. y Tsay, R. S. (2000) "Detection of Outlier Patches in Autoregressive Time Series", *Statistica Sinica*, **11**, 651-673.

- Kaiser, R. y Maravall, A. (2002) "Seasonal Outliers in Time Series", *Estadística*, **53**.
- Kapetanios, G. (2001) "Model Selection in Threshold Models", *Journal of Time Series Analysis*, **22**, 733-754.
- Kass, R. E. y Raftery A. E. (1995) "Bayes Factors", *Journal of the American Statistical Association*, **90**, 773-795.
- Kass, R. E. y Wasserman, L. A. (1995) "A Reference Bayesian Test for Nested Hypothesis and Its Relationship to the Schwarz Criterion", *Journal of the American Statistical Association*, **90**, 928-934.
- Kavalieris, L. (1989) "The Estimation of the Order of an Autoregression Using Recursive Residuals and Cross-validation", *Journal of Time Series Analysis*, **10**, 271-281.
- Kim, S., Cho, S. y Lee S. (2000) "On the Cusum Test for Parameter Changes in GARCH(1,1) Models", *Communications in Statistics: Theory and Methods*, **29**, 115-162.
- Kim, T., Leybourne, S. y Newbold, P. (2002) "Unit Root Tests with a Break in Innovation Variance", *Journal of Econometrics*, **109**, 365-387.
- Kitagawa, G. (1979) "On the Use of AIC for the Detection of Outliers", *Technometrics*, **21**, 193-199.
- Konishi, S. y Kitagawa, G. (1996) "Generalised Information Criteria in Model Selection", *Biometrika*, **83**, 875-890.
- Koreisha, S. y Pukkila, T. (1993) "Determining the Order of a Vector Autoregression When the Number of Component Series is Large", *Journal of Time Series Analysis*, **14**, 47-69.
- Le, N. D., Martin, R. D. y Raftery, A. E. (1996) "Modeling Flat Stretches, Bursts, and Outliers in Time Series Using Mixture Transition Distribution Models", *Journal of the American Statistical Association*, **91**, 1504-1515.
- Ledolter, J. (1989) "The Effect of Additive Outliers on the Forecasts From ARIMA Models", *International Journal of Forecasting*, **5**, 231-240.
- Ledolter, J. (1990) "Outlier Diagnostics in Time Series Analysis", *Journal of Time Series Analysis*, **11**, 317-324.
- Lee, C. H. y Martin, R. D. (1982) "M-estimates for ARMA Processes", Technical Report, University of Washington.
- Lee, S. y Park, S. (2001) "The Cusum of Squares Test for Scale Changes in Infinite Order Moving Average Models", *Scandinavian Journal of Statistics*, **28**, 625-644.
- Lee, S., Tokutsu, Y. y Maekawa, K. (2003) "The Residual Cusum Test for the Constancy Parameters in GARCH(1,1) Models", Manuscrito.

- Leeuw, J. van der (1994) "The Covariance Matrix of ARMA Errors in Closed Form", *Journal of Econometrics*, **63**, 397-405.
- Lefrançois, B. (1991) "Detecting Over-influential Observations in Time Series", *Biometrika*, **78**, 91-99.
- Ljung, G. (1993) "On Outlier Detection in Time Series", *Journal of the Royal Statistical Society B*, **55**, 559-567.
- Luceño, A. (1998) "Detecting Possibly Non-Consecutive Outliers in Industrial Time Series", *Journal of the Royal Statistical Society B*, **60**, 295-310.
- Lütkepohl, H. (1985) "Comparison of Criteria for Estimating the Order of a Vector Autoregressive Process", *Journal of Time Series Analysis*, **6**, 35-52.
- Lütkepohl, H. (1993) *Introduction to Multiple Time Series Analysis*, 2nd Ed., Springer-Verlag: New York.
- Lütkepohl, H. y Poskitt, D. S. (1996) "Specification of Echelon-Form VARMA models", *Journal of the Business and Economic Statistics*, **14**, 69-79.
- Maravall, A. y Mathis, A. (1994) "Encompassing Univariate Models in Multivariate Time Series", *Journal of Econometrics*, **61**, 197-233.
- Martin, R. D. (1980) "Robust Estimation for Autoregressive Models", *Directions in Time Series*, Institute of Mathematical Statistics, 228-254.
- Martin, R. D. (1981) "Robust Methods in Time Series", *Applied Time Series Analysis II*, Academic Press, 683-759.
- Martin, R. D., Samarov, A. y Vandalae, W. (1983) "Robust Methods for ARIMA Models", *Applied Time Series Analysis of Economic Data*, Bureau of Census, 153-169.
- McCulloch, R. E. y Tsay, R. S. (1993) "Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series", *Journal of the American Statistical Association*, **88**, 968-978.
- McCulloch, R. E. y Tsay, R. S. (1994) "Bayesian Analysis of Autoregressive Time Series Via the Gibbs Sampler", *Journal of Time Series Analysis*, **15**, 235-250.
- McQuarrie, A., Shumway, R. y Tsai, C. (1997) "The Model Selection Criterion AICu", *Statistics and Probability Letters*, **34**, 285-292.
- Menzeffricke, U. (1981) "A Bayesian Analysis of a Change in the Precision of a Sequence of Independent Normal Random Variables at an Unknown Time Point", *Applied Statistics*, **30**, 141-146.

- Nelson, D. B. (1991) "Conditional Heteroskedasticity in Asset Returns: A New Approach", *Econometrica*, **59**, 347-370.
- Ogata, Y. (1980) "Maximum Likelihood Estimates of Incorrect Markov Models for Time Series and the Derivation of AIC", *Journal of Applied Probability*, **17**, 59-72.
- Park, S., Lee, S. y Jeon J. (2000) "The Cusum of Squares Test for Variance Changes in Infinite Order Autoregressive Models", *Journal of the Korean Statistical Society*, **29**, 351-361.
- Parzen, E. (1974) "Some Recent Advances in Time Series Modelling", *IEEE Transactions on Automatic Control*, **19**, 723-730.
- Parzen, E. (1975) "Some Solutions to the Time Series Modelling and Prediction Problem", *The Search for Oil*. Marcel Dekker: New York, 1-16.
- Paulsen, J. (1984) "Order Determination of Multivariate Autoregressive Time Series With Unit Roots", *Journal of Time Series Analysis*, **5**, 115-127.
- Peña, D. (1987) "Measuring the Importance of Outliers in ARIMA Models. New Perspectives in Theoretical and Applied Statistics", John Wiley: New York, 109-118.
- Peña, D. (1990) "Influential Observations in Time Series", *Journal of the Business and Economic Statistics*, **8**, 235-241.
- Peña, D. (1991) "Measuring Influence in Dynamic Regression Models", *Technometrics*, **33**, 93-101.
- Peña, D. y Box, G. E. P. (1987) "Identifying a Simplifying Structure in Time Series", *Journal of the American Statistical Association*, **82**, 836-843.
- Peña, D. y Prieto, F. J. (2001, a) "Multivariate Outlier Detection and Robust Covariance Matrix Estimation (with discussion)", *Technometrics*, **43**, 286-310.
- Peña, D. y Prieto, F. J. (2001, b) "Cluster Identification Using Projections", *Journal of the American Statistical Association*, **96**, 1433-1445.
- Peña, D. y Sánchez, I. (2004) "Multifold Predictive Validation in ARMAX Time Series Models", *Journal of the American Statistical Association*, To appear.
- Peña, D., Tiao, G. C. y Tsay, R. S. (2000) *A Course in Time Series Analysis*. John Wiley & Sons: New York.
- Poskitt, D. S. (1987) "A Modified Hannan-Rissanen Strategy for Mixed Autoregressive-Moving Average Order Determination", *Biometrika*, **74**, 781-790.
- Poskitt, D. S. (1989) "A Two Stage Least Squares Procedure for the Identification of AR-MAX Models", *Journal of the Royal Statistical Society B*, **51**, 29-46.



- Posse, C. (1995) "Tools for Two-dimensional Exploratory Projection Pursuit", *Journal of Computational and Graphics Statistics*, **4**, 83-100.
- Pötscher, B. M. (1990) "Estimation of Autoregressive Moving-Average Order Given an Infinite Number of Models and Approximation of Spectral Densities", *Journal of Time Series Analysis*, **11**, 165-179.
- Priestley, M. B. (1981) *Spectral Analysis and Time Series*, London: Academic Press.
- Pukkila T., Koreisha, S. y Kallinen, A. (1990) "The Identification of ARMA Models", *Biometrika*, **77**, 537-548.
- Quinn, B. G. (1980) "Order Determination for a Multivariate Autoregression", *Journal of the Royal Statistical Society B*, **42**, 182-185.
- Raftery, A. E., Madigan, D. y Volinsky, C. T. (1996) "Accounting for Model Uncertainty in Survival Analysis Improves Predictive", *Bayesian Statistics 5*, Oxford University Press: Oxford, 323-349.
- Ramsey, F.L. (1974) "Characterization of the Partial Autocorrelation Function", *Annals of Statistics*, **2**, 1296-1301.
- Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*, 2nd Ed. John Wiley & Sons: New York.
- Reinsel, M. B. (1997) *Elements of Multivariate Time Series*, 2nd Ed. New York: Springer-Verlag.
- Rissanen, J. (1978) "Modelling by Shortest Data Description", *Automatica*, **14**, 465-471.
- Sánchez, M. J. y Peña, D. (2003) "The Identification of Multiple Outliers in ARIMA Models", *Communications in Statistics: Theory and Methods*, **32**, 1265-1287.
- Sansó, A., Aragó, V. y Carrion, J. L. (2003) "Testing for Changes in the Unconditional Variance of Financial Time Series", Working Paper, Universidad de Baleares.
- Schwarz, G. (1978) "Estimating the Dimension of a Model", *Annals of Statistics*, **6**, 461-464.
- Sensier, M. y Dijk, D. V. (2004) "Testing for Volatility Changes in US Macroeconomic Time Series", *Review of Economics and Statistics*, To appear.
- Shao, J. (1996) "Bootstrap Model Selection", *Journal of the American Statistical Association*, **91**, 655-665.
- Shi, P. y Tsai, C. (1998) "A Note on the Unification of the Akaike Information Criterion", *Journal of the Royal Statistical Society B*, **60**, 551-558.

- Shi, P. y Tsai, C. (2002) "Regression Model Selection - A Residual Likelihood Approach", *Journal of the Royal Statistical Society B*, **64**, 237-252.
- Shibata, R. (1976) "Selection of the Order of an Autoregressive Model by Akaike's Information Criterion", *Biometrika*, **63**, 117-126.
- Shibata, R. (1980) "Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process", *Annals of Statistics*, **8**, 147-164.
- Shibata, R. (1989) "Statistical Aspects of Model Selection", *From Data to Model*. Springer, 215-240.
- Shibata, R. (1997) "Bootstrap Estimate of Kullback-Leibler Information for Model Selection", *Statistica Sinica*, **7**, 375-394.
- Shumway, R. H. y Stoffer, D. S. (2000) *Time Series Analysis and Its Applications*. Springer Verlag: New York.
- Smith, A. F. M. (1975) "Bayesian Approach to Inference About a Change-Point in a Sequence of Random Variables", *Biometrika*, **62**, 407-416.
- Söderstrom, T. (1977) "On Model Structure Testing in Systems Identification", *International Journal of Control*, **26**, 1-18.
- Stigler, S. M. (1986) *The History of Statistics*. Harvard University Press: Cambridge.
- Stoica, P., Eykhoff, P., Janssen, P. y Söderström, T. (1986) "Model-Structure Selection by Cross-Validation", *International Journal of Control* **43**, 1841-1878.
- Stone, M. (1977) "An Asymptotic Equivalence of Choice Of Model by Cross-validation and Akaike's Criterion", *Journal of the Royal Statistical Society B*, **39**, 44-47.
- Stone, C. J. (1982) "Local Asymptotic Admissibility of a Generalization of Akaike's Model Selection Rule", *Annals of the Institute of Statistical Mathematics*, **34**, 123-133.
- Sugiura, N. (1978) "Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections", *Communications in Statistics: Theory and Methods*, **7**, 13-26.
- Taylor, S. J. (1986) *Modelling Financial Time Series*. John Wiley: New York.
- Tiao, G. C. y Tsay, R. S. (1989) "Model Specification in Multivariate Time Series (with discussion)", *Journal of the Royal Statistical Society B*, **51**, 157-213.
- Tierney, L. y Kadane, J. B. (1986) "Accurate Approximations for Posterior Moments and Marginal Densities", *Journal of the American Statistical Association*, **81**, 82-86.
- Tong, H. (1979) "A Note on the Local Equivalence of Two Recent Approaches to Autoregressive Order Determination", *International Journal of Control*, **29**, 441-446.

- Tong, H. (1983) *Threshold Models in Non-linear Time Series Analysis*. Springer Verlag: New York.
- Tong, H. (1990) *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press: Oxford.
- Trívez, F. J. (1995) "Level Shifts, Temporary Changes and Forecasting", *Journal of Forecasting*, **14**, 543-550.
- Trívez, F. J. y Nievas, J. (1998) "Analyzing the Effects of Level Shifts and Temporary Changes on the Identification of ARIMA Models", *Journal of Applied Statistics*, **25**, 409-424.
- Tsay, R. S. (1984) "Order Selection in Nonstationary Autoregressive Models", *Annals of Statistics*, **12**, 1425-1433.
- Tsay, R. S. (1986) "Time Series Model Specification in the Presence of Outliers", *Journal of the American Statistical Association*, **81**, 132-141.
- Tsay, R. S. (1987) "Conditional Heteroscedastic Time Series Models", *Journal of the American Statistical Association*, **82**, 590-604.
- Tsay, R. S. (1988) "Outliers, Level Shifts and Variance Changes in Time Series", *Journal of Forecasting*, **7**, 1-20.
- Tsay, R. S. (2002) *Analysis of Financial Time Series: Financial Econometrics*. John Wiley & Sons: New York.
- Tsay, R. S., Peña, D. y Pankratz, A. E. (2000) "Outliers in Multivariate Time Series", *Biometrika*, **87**, 789-804.
- Wang, J. y Zivot, E. (2001) "A Bayesian Time Series Model of Multiple Structural Changes in Level, Trend, and Variance", *Journal of the Business and Economic Statistics*, **18**, 374-386.
- Wichern, D. W., Miller, R. B. y Hsu, D. A. (1976) "Changes of Variance in First Order Autoregressive Time Series Models - With An Application", *Applied Statistics*, **25**, 248-256.
- Wong, C. S. y Li, W. K. (1998) "A Note on the Corrected Akaike Information Criterion for the Threshold Autoregressive Models", *Journal of Time Series Analysis*, **19**, 113-124.
- Wu, L. S., Hosking, J. R. M. y Ravinshanker, N. (1993) "Reallocation Outliers in Time Series", *Applied Statistics*, **42**, 301-313.
- Zellner, A. (1987) *An Introduction to Bayesian Inference in Econometrics*. John Wiley & Sons: New York.

- Zhang, P. (1993) "Model Selection Via Multifold Cross-Validation", *Annals of Statistics*, **21**, 299-313.